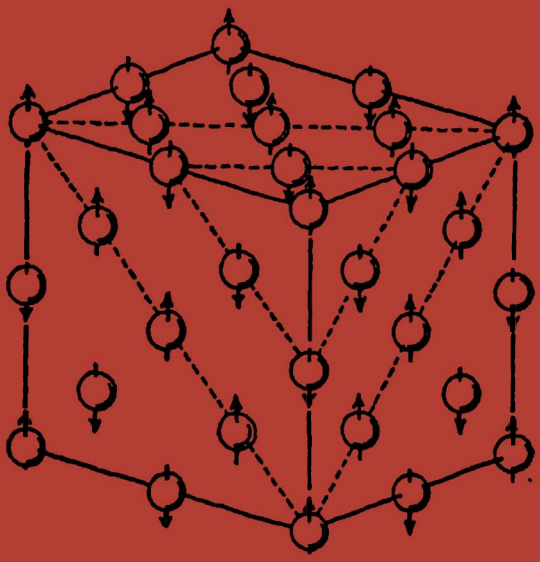
A.F. IOFFE

PHYSICS OF SEMI-CONDUCTORS



PHYSICS of SEMICONDUCTORS

A. F. IOFFE

Member of the U.S.S.R. Academy of Sciences,
Director of the Institute of Semiconductors

ACADEMIC PRESS INC.
NEW YORK

Translated from Russian

A. F. IOFFE
"Fizika Poluprovodnikov", 491 pp., 1957.

Published by the Publishing House of the U.S.S.R. Academy of Sciences, Moscow–Leningrad. Revised and supplemented by Prof. A. F. Ioffe for the English edition. Translation edited and indexed by H. J. Goldsmid.

© 1960 by Infosearch Ltd.

All Rights Reserved.

This book or any part thereof must not be reproduced without the written permisson of the publishers

Infosearch Ltd.
207 Brondesbury Park,
London, N.W.2.

In Western Hemisphere

Academic Press Inc.
111 Fifth Avenue
New York 3, N.Y.

PRINTED IN GREAT BRITAIN BY JOHN WRIGHT & SONS LTD., AT THE STONEBRIDGE PRESS, BRISTOL

Contents

List of Symbols . . . ix

Chapter 1. Solid Electrolytes . . . 1
- Movement of charge in electrolytes . . . 1
- The nature of the current carriers . . . 2
- The current mechanism in solid electrolytes . . . 5
- Concentration of charge carriers . . . 9
- Temperature dependence of the conductivity . . . 11
- Ion mobility . . . 13
- Crystal lattice defects . . . 15
- Impurities . . . 16
- Chemical nature of the current carriers in ionic crystals . . . 17
- Mixed ionic and electronic currents . . . 19
- High-voltage polarization . . . 20
- The effect of ionic currents in electronic conductors . . . 22

Chapter 2. Metals . . . 24
- The motion of electrons . . . 24
- Electron mobility . . . 25
- The effect of temperature on electron mobility . . . 26
- Thermal conductivity . . . 27
- Temperature dependence of the conductivity . . . 30
- Electron theory of metals . . . 32
- Quantum statistics of electrons in metals . . . 34
- Velocity distribution of electrons . . . 35
- The effect of temperature. . . . 37
- Energy distribution of electrons . . . 38
- Degeneracy of the electron gas . . . 41
- Heat capacity of electrons . . . 42
- Quantum states of the electrons in an atom . . . 43

EC 2 1 '61

Quantum states of electrons in a metal 44
X-ray spectra of metals 45
Distribution of electron velocities in an electrical field . . . 49
Thermionic emission 51
Magnetic properties of electrons 55
Thermoelectric phenomena 56

CHAPTER 3. ELECTRONIC SEMICONDUCTORS— A GENERAL SURVEY . . 59
Types of semiconductor 59
Comparison with electrolytes 60
Comparison with metals 61
Sign of the current carriers 63
Thermal equilibrium of the electrons in a semiconductor . . . 66
Temperature-dependence of the electron concentration and electrical conductivity 68
The effects of impurities on the conduction mechanism and electrical conductivity 69
Photoconduction 72
Volume photo-e.m.f.s. 79
Strong electric fields and breakdown 79
The barrier layer 82
The need for the quantum theory in studies on semiconductors . 84

CHAPTER 4. FUNDAMENTALS OF THE QUANTUM THEORY OF SEMICONDUCTORS 87
The quantum states of electrons in a solid 87
Conduction electrons 91
The laws of motion for free electrons 96
The conditions of motion of an electron in a semiconductor . 97
The laws of motion of electrons in a periodic field . . . 99
The effective mass 100
The Brillouin zones 104
Holes and electrons 104
Carrier mobilities 109
Methods of measuring mobilities and effective masses . . . 110
Additions to the electron theory of metals 111
The distribution of the electrons in a semiconductor among the quantum states 112
The free electron and hole concentrations 114

The chemical potential 116
The X-ray spectra of semiconductors . . 118
Semiconductors of low resistivity . . . 119
Impurity levels 120
Surface states 125
Polarons 127
Excitons 128
The energy levels of semiconductors . . 129
The present state of semiconductor theory . . 132

CHAPTER 5. PHYSICAL PHENOMENA IN SEMICONDUCTORS (PROPERTIES OF THE CRYSTAL LATTICE) 136
Short- and long-range order 136
Semiconductor crystal lattices 139
Imperfections in crystal lattices 148
Impurities 151
Thermal vibrations of the lattice 156
Scattering of electrons 167
Mean free path and mobility of electrons . . 173
The concentration of charge carriers 184
Properties of surface layers on semiconductors 188
Contact resistance 189
Solid rectifiers 196
Slow time-dependent phenomena in barrier layers 203
p–n junctions 205
Rectifiers and transistors 214
Transformation of light and nuclear radiation into electrical energy by means of *p–n* junctions 219
Emission of charge carriers from a semiconductor 221
The contact potential 228
External photo-effect 231
Thermionic emission 234
Secondary electron emission 237
Electron emission due to ions 240
Contact electrification 242
Semiconductors in electric and magnetic fields 243
Change of resistance in a magnetic field 249
Diamagnetism 257

Paramagnetism 258
Cyclotron or diamagnetic resonance 262
Ferromagnetism 266
Ferrites 270
Thermal conductivity of semiconductors 271
Thermoelectricity: discovery of thermoelectric phenomena . . 288
Calculation of the thermal e.m.f. 305
Thermoelectric generators 313
Thermoelectric cooling 318
Thermoelectric heating 321

CHAPTER 6. METHODS OF MEASURING THE BASIC PROPERTIES OF SEMICONDUCTORS 326
Electrical conductivity measurements 328
Mobility 331
Carrier concentration and effective mass 332
Forbidden band width and position of impurity levels . . . 333
The lifetime and the diffusion length 335
The thermoelectric power 336
Thermal conductivity and heat capacity 337
The contact potential 342
Photoconductivity 344
The dielectric constant 346

CHAPTER 7. SEMICONDUCTING MATERIALS 349
The study of semiconducting materials 349
The classification of semiconductors 352
Atomic lattices with valence bonds 353
Molecular lattices 376
Sulphides, selenides and tellurides 381
Oxides 386
Ionic crystals 395
Alloys of metals 396
Liquid semiconductors 399
Organic materials 400
Semiconductor materials for thermoelements 401

CONCLUSIONS 423

SUBJECT INDEX 433

List of Symbols

In the following list the various symbols are defined as they are most commonly employed in the book. However, in some cases it will be found that a symbol may have a meaning differing from that given below; in all such cases the interpretation of the symbol will be clear from the context.

A constant in Hall formula,
C constant in thermoelectric equation, constant of interaction,
D diffusion coefficient,
D_e diffusion coefficient of current carriers,
E energy,
$\mathscr{E}$ electric field,
ΔE_0 energy gap,
ΔE impurity activation energy,
$\Delta E'$ distance of Fermi level from band edge,
E_1 energy at top of valence band,
E_2 energy at bottom of conduction band,
E_ϕ energy at Fermi level,
F free energy, force,
H magnetic field,
H_c coercive force,
I angular momentum,
J electric current,
K thermal diffusivity,
L length, diffusion length, barrier layer thickness, light intensity,
M ionic or atomic mass, magnetic moment,
N excess impurity concentration, density of ions, Nernst coefficient,
N_0 number of atoms per cm^3,
N_1 donor concentration,
N_2 acceptor concentration,
N_{eff} effective density of states,
P momentum,

Q quantity of electrons, quantity of heat,
R Hall coefficient, resistance,
S cross-section area, entropy, Righi–Leduc coefficient,
T absolute temperature,
T_k Fermi temperature,
T_K Curie temperature,
T_1 hot junction temperature,
T_2 cold junction temperature,
U energy, ionization energy,
V volume, voltage or potential difference,
V_c contact potential difference,
V_n polarization voltage,
a lattice constant,
c specific heat, velocity of light,
c' thermal capacity per unit volume,
e electronic charge,
f perturbed Fermi function,
f_0 equilibrium Fermi function,
g spectroscopic splitting factor,
h Planck's constant,
j current density,
k optical absorption coefficient, Boltzmann's constant, wave number,
l free path length,
m mass of carrier,
m_0 free electron mass,
m_-^* effective mass of electrons,
m_+^* effective mass of holes,
n refractive index, carrier density,
n_0 intrinsic electron concentration,
n_- electron concentration,
n_+ hole concentration,
p proportion of filled states, pressure, exponent in temperature variation of mobility,
$p_{1,2,3}$ momentum coordinates,
q rate of heating, charge,
$q_{1,2,3}$ space coordinates,
r exponent of energy in expression for free path length,
t time,
u mobility,
v electron velocity, velocity of sound,
x absorption coefficient,
z thermoelectric figure of merit,

α thermoelectric power, temperature coefficient of energy gap,
γ coefficient related to polarizability,
ϵ dissociation energy of ion, dielectric constant, kinetic energy of electron,
ζ coefficient of performance,
η efficiency,
θ Debye temperature,
κ thermal conductivity,
κ_1 lattice component of thermal conductivity,
κ_2 electronic component of thermal conductivity,
λ wavelength, $-\mu/kT$,
$\bar{\lambda}$ mean free path of phonons,
μ chemical or Fermi potential, Bohr magneton, permeability,
ν frequency of vibration,
Π Peltier coefficient,
ρ electrical resistivity, space charge density, density,
σ electrical conductivity, secondary emission coefficient,
σ_p photoconductivity,
τ relaxation time, Thomson coefficient,
$\bar{\tau}$ electron or hole relaxation time,
τ_0 lifetime,
ϕ work function,
χ magnetic susceptibility,
ψ wave function,
ω barrier energy.

CHAPTER

1

Solid Electrolytes

Movement of charge in electrolytes

The mechanism of electrolytic conduction in aqueous solutions was thoroughly investigated and understood even in the 19th century. A similar mechanism was also established later for currents in many solid salts and their melts. It was found that currents in most glasses, in polymers and many insulating materials are also carried by ions.

Electric currents observed in such solids differ in many respects from currents passing through liquid solutions, despite the similarity of the charge carriers; in the solids the magnitude of the current for a constant potential difference usually decreases with time and often drops virtually to zero. When the source of the potential difference is removed, one observes appreciable reverse e.m.f.s, in closed circuits of many solid electrolytes, which attain values equal to the earlier applied potential difference; reverse currents appear even without the application of an external e.m.f. Sometimes these reverse currents transfer a quantity of electricity equal to that which has previously passed in the forward direction. Heat and light alter the electrical properties of such conductors, the changes sometimes being irreversible. These properties of dielectric crystals were discovered in the 19th century and studied in detail by Paul Curie.

As late as 1912 the foregoing phenomena were still regarded as abnormal manifestations of dielectric properties of insulators and not as a result of the movement of charges in the materials.

I have, however, succeeded in demonstrating that the process does not involve the dielectric displacement of charge within a molecule but the movement of charge across the bulk of the conductor, i.e. the flow of an electric current. The special features of the current in solid electrolytes are due to the accumulation of charges of one sign in the vicinity of the electrodes. The field set up by such a space charge is superimposed on the external field, thus leading to a change in the magnitude of the current.

When the space charge is duly taken into account the apparent anomalies disappear. The current obeys Ohm's law over a wide range

of potential differences, the current density being strictly proportional to the field strength. Solid electrolytes can therefore be described not only by their dielectric constant but also by a certain specific electrical conductivity which depends strongly on temperature and chemical composition, this dependence following certain definite laws.

My colleagues and I have succeeded in showing that the extensive scatter in the values of the electrical conductivity of individual specimens of ionic crystals may be attributed to different impurity contents. By repeated crystallization we prepared some tens of pure specimens of alum and of saltpetre which respectively had identical electrical conductivities and temperature dependences of the electrical conductivity.

The specific electrical conductivity σ can always be expressed as a product of the concentration n of current carriers, their charge e and their mobility u (defined as the average translation velocity along unit electric field)

$$\sigma = neu. \tag{1}$$

The current density j, equal to the ratio of the current J to the cross-section area S of the conductor, is given by

$$j = neu\mathscr{E} \tag{2}$$

where $\mathscr{E}$ is the electric field.

A radioactive tracer technique now makes it possible to observe the average mobility of ion transfer along the field directly and thus to separate n and u in expression (1).

The mobility of ions in an electrical field can also be estimated from the diffusion coefficient D. Einstein established a universal relationship between D and u

$$\frac{D}{u} = \frac{kT}{e} \tag{3}$$

where k is Boltzmann's constant which is equal to $1{\cdot}38 \times 10^{-16}$ erg/degree. Most measurements of ion mobility were, however, carried out before radioactive tracers made their appearance, using indirect chemical methods.

The nature of the current carriers

The electrolytic nature of the current can be established by the liberation of the products of electrolysis at the electrodes, in quantities prescribed by Faraday's law. Errors may, however, occur due to secondary processes at the electrodes or at the boundary between two solid electrolytes forming an electrical circuit. The metal deposited at the cathode often forms sharp needles, dendrites, of low resistivity, penetrating into the crystal. These internal metallic needles become electrodes at which the products of electrolysis are liberated. This may

lead to an appreciable source of error in the subsequent weighing of the electrode. When such needles reach the anode they become metallic bridges carrying the current and the electrolysis of the crystals ceases. It is, therefore, essential to eliminate the formation of dendrites at the cathode when carrying out such investigations.

The proof of the ionic character of the current becomes more convincing when the proportions of current transferred by each type of ion, both positive and negative, are measured. For this purpose the conductor is divided into three or more parts, each of which is weighed separately before the experiment and after the passage of current J during time t. The end portions are weighed together with the electrodes (Fig. 1).

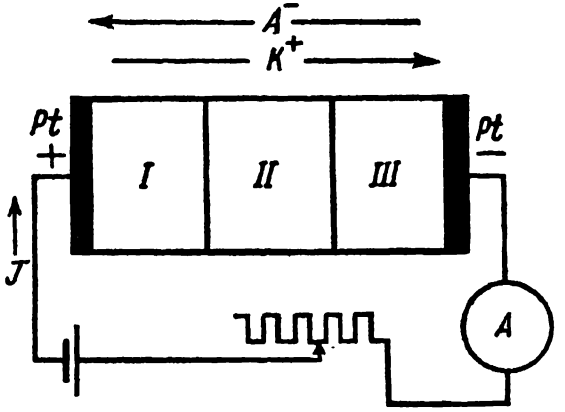


Fig. 1. An experiment for determining transfer numbers in solid electrolytes: A^-, anions; K^+, cations; I, II, III, parts of the conductor.

When a quantity of electricity Q equal to Jt passes through the conductor, a corresponding amount of the substance is transferred by the ions. By determining the change of the weight of each of the three portions it is possible to estimate the amount of substance liberated by the current at each electrode and the proportion of positive and negative ions in the total current. Thus, for example, when the current is transferred only by positive ions an increase in weight corresponding to Faraday's law is observed in the portion adjoining the cathode which is equal to the loss in weight of the portion adjoining the anode.

Verification of Faraday's law and measurement of the ion transfer numbers can only be carried out with appreciable currents and, therefore, with a reasonably high electrical conductivity of the material. At low electrical conductivities, one has to be satisfied with indirect

symptoms in order to decide whether the current mechanism is ionic or electronic. The following criteria can be used:

1. Ionic conduction is indicated by electrical polarization at the electrodes which keeps the current virtually equal to zero up to a certain potential difference V_0 of the order of 1–2 V; when this is exceeded the current is proportional to $V - V_0$.

2. E.m.f.s may be set up in a conductor carrying a current in a magnetic field. When the magnetic field is perpendicular to the direction of the current, an additional electrical field perpendicular to the magnetic and electrical fields appears in electronic conductors (a transverse potential difference). When the current mechanism is ionic this phenomenon (the Hall effect) is not observable. Even in electronic conductors this field is sometimes so weak that it remains unnoticed.[1]

3. The temperature dependence of the electrical conductivity makes it possible to estimate the energy required for the liberation and movement of the charges in the conductor. When this energy is the same at all temperatures and the electrolytic nature of the current has been established at high temperatures by the applicability of Faraday's law and the measurement of ion transfer numbers, then the current at low temperatures may also be attributed to ions. As a rule, however, the same dissociation energy is not applicable at high and low temperatures.

Cases of mixed ionic and electronic conductivity are not infrequent. Strict distinction between the two mechanisms often represents a complex problem requiring a thorough investigation of the behaviour of the materials over a wide temperature range.

How difficult it may sometimes be to establish the true character of current carriers is indicated by the example of the high-temperature modification of silver sulphide, Ag_2S.

In 1833, Faraday, studying the currents in Ag_2S, discovered that, in contrast to the behaviour of a metal, its electrical conductivity rises with temperature reaching values characteristic of the majority of metals at high temperatures. Twenty years later Hittorf reached the conclusion that α-Ag_2S is a solid electrolyte. Fifty years later α-Ag_2S, together with other sulphides, was classified among metallic conductors.[2]

In 1920 Tubandt measured the Faraday constant and the transfer numbers for α-Ag_2S and found that all the current is transferred by silver ions. It seemed that the nature of conductivity of α-Ag_2S was definitely established.

There were, however, several signs pointing towards an electronic conduction mechanism:

1. The calculated ion mobilities based on Tubandt's hypothesis gave values exceeding those for all known solid electrolytes. The diffusion coefficient D was measured for α-Ag_2S and related to the mobility u by

Einstein's law (expression (3)). The mobility calculated on the basis of the measured diffusion coefficient was 1000 times less than expected.

2. Klaiber observed an appreciable Hall effect which is a sure sign of electronic conduction.

3. The e.m.f. of a galvanic element made of platinum, sulphur, α-Ag_2S and silver was found to be equal to a few thousandths of a volt as compared with 0·2 V expected from an electrolytic cell.

Finally, in 1933 Wagner proved that an error had crept into Tubandt's experiments; the liberation of silver and sulphur was not due to electrolysis but to diffusion of silver atoms through silver sulphide, the atoms forming silver sulphide at the sulphur boundary.

One can now consider it to be definitely established that the currents in the high-temperature modification of silver sulphide are electronic and the part played by ionic conduction, if it takes place at all, does not exceed 1%.

The current mechanism in solid electrolytes

At low temperatures each ion is fixed in the crystal lattice in a position corresponding to the minimum potential energy. In order to transfer the ion from this position into a new one so that it can move in the direction of the electrical field it is necessary to expend a certain amount of energy, called the dissociation energy ϵ. This energy may be provided by an appropriate fluctuation of its thermal motion. When an ion may exist in one of two states, the energies of which differ by ϵ, the ratio of the time it stays in the higher energy state to that in the lower energy state is given by the expression $\exp(-\epsilon/kT)$, where k is the Boltzmann's constant and T the absolute temperature.

Denoting by N the total number of ions per unit volume which may dissociate with an energy change ϵ, and the equilibrium number of dissociated ions which can participate in electrical conduction by N_i, we can calculate N_i approximately in the following way.

The number of ions which dissociate per unit time is

$$p = \alpha N \exp(-\epsilon/kT). \tag{4}$$

The number of ions which recombine in unit time is

$$z = \beta N_i \frac{N_i}{N}. \tag{5}$$

From the equilibrium condition we have $p = z$, whence

$$N_i = \sqrt{\left(\frac{\alpha}{\beta}\right)} N \exp(-\epsilon/2kT). \tag{6}$$

The dissociated ions move from positions, corresponding to the minimum of their free energy in the crystal lattice, into new positions (e.g. into spaces in the crystal cell between adjacent ions). The potential energy there is ϵ eV higher than the energy which they possess when occupying a lattice site (Fig. 2).

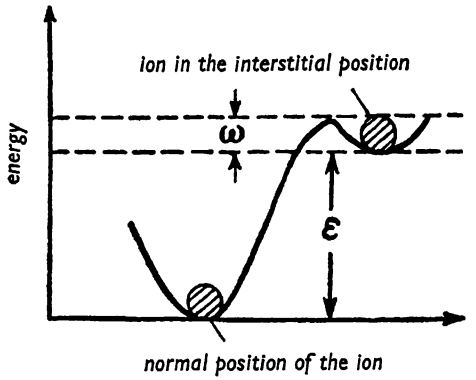


Fig. 2. Change of the potential energy of an ion on transition into an interstitial site of the crystal lattice (formation of defects according to Frenkel').

Similar positions with the same energy also exist in adjacent cells. In order to pass from one cell into another and to contribute in this way to the current, the ion must overcome a certain energy barrier each time, penetrating structural elements of the solid and then passing through intermediate positions where its energy will be higher by ω eV. This additional energy is picked up by the ion from the thermal vibrations.

Under the effect of thermal agitation a dissociated interstitial ion undergoes a continuous oscillation with a frequency ν depending on the forces acting on it. The ion therefore approaches the barrier, of height ω eV, ν times per sec. The probability that in one of these attempts it will succeed in overcoming the barrier, raising its energy by ω eV, is given by the expression derived from statistical physics

$$\exp(-\omega/kT). \tag{7}$$

It is therefore expected that in unit time the ion will pass $\nu \exp(-\omega/kT)$ times from cell to cell, diffusing through the crystal until it reaches a vacant lattice site where an ion of the same sign is absent, or until it attaches itself to a crystal defect with a lower energy than that on an interstitial site.

The energy ω received by the ion in overcoming the barrier is carried over in the form of increased oscillations as the ion passes into

a new cell. When the amplitude of the oscillation decreases, the ion gives up its surplus energy ω to the surrounding atoms in the form of thermal energy.

When ω is small in comparison with ϵ, an ion which has left the equilibrium position at the lattice site will diffuse from one cell into another, moving equally often to the left and to the right, forwards and backwards, and upwards and downwards.

When an electrical field $\mathscr{E}$ directed e.g. from left to right, exists in the crystal, all the positive ions will begin to move more frequently in this direction than in the opposite direction, whilst the negative ions will be assisted by the field in moving from the right to the left. It may be assumed that the predominance of transition in the direction in which the field acts over the opposite direction will on an average be proportional to the field strength $\mathscr{E}$. The average distance covered by an ion in one second in the direction of the field or its average velocity may be expressed as a product of the total number of transitions $\nu \exp(-\omega/kT)$ and the probability of transition in the direction of the field, which is proportional to $\mathscr{E}$, and will be denoted by $\gamma\mathscr{E}$. We have, therefore,

$$v = \gamma\mathscr{E}\nu \exp(-\omega/kT)$$

and the mobility u_i, or average translation velocity in unit field is

$$u_i = \gamma\nu \exp(-\omega/kT).$$

The electrical conductivity σ, according to expression (1) can be written either as

$$\sigma_i = e\frac{\alpha\gamma\nu N_i}{\beta_1}\exp\left(-\frac{\omega+\epsilon}{kT}\right), \tag{8}$$

when there is a large number of defects to which the ions may attach themselves, or as

$$\sigma_i = e\gamma\nu\sqrt{\left(\frac{\alpha}{\beta_2}\right)}N_i^{\frac{1}{2}}\exp\left(-\frac{\epsilon+2\omega}{2kT}\right), \tag{8a}$$

when the ions recombine with empty lattice sites.

After the crystal has been maintained at temperature T for a sufficiently long time, an equilibrium will be established under which the ions will become distributed in the most probable way between their (a) normal positions in crystal lattice sites, (b) cracks and defects, and (c) interstitial sites with increased energy ϵ. Similarly, a stationary distribution independent of time will be set up in an electric field $\mathscr{E}$. As long as Ohm's law is valid it ought to be considered that the electrical field $\mathscr{E}$ does not appreciably change the equilibrium distribution and, therefore, the concentration of the free ions.

Often, when it is not possible to separate the processes governing the concentration of free charges N_i from processes governing their mobility u_i, σ_i is expressed by a formula combining the two effects (8 and 8a)

$$\sigma_i = \sigma_0 \exp\left(-\frac{E_0}{kT}\right). \tag{8b}$$

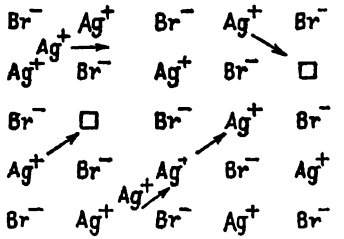


Fig. 3. Schematic representation of defects in a crystal lattice according to Frenkel'. Ag^+ positive ions occupy the interstices of the crystal lattice AgBr. There is an equal number of vacant sites in the cationic partial lattice.

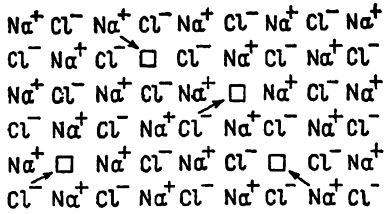

Na⁺ Cl⁻ Na⁺ Cl⁻ Na⁺ Cl⁻ Na⁺ Cl⁻ Na⁺
Cl⁻ Na⁺ Cl⁻ □ Cl⁻ Na⁺ Cl⁻ Na⁺ Cl⁻
Na⁺ Cl⁻ Na⁺ Cl⁻ Na⁺ □ Na⁺ Cl⁻ Na⁺
Cl⁻ Na⁺ Cl⁻ Na⁺ Cl⁻ Na⁺ Cl⁻ Na⁺ Cl⁻
Na⁺ □ Na⁺ Cl⁻ Na⁺ Cl⁻ □ Cl⁻ Na⁺
Cl⁻ Na⁺ Cl⁻ Na⁺ Cl⁻ Na⁺ Cl⁻ Na⁺ Cl⁻

Fig. 4. Schematic representation of defects in the crystal lattice of rock salt, according to Schottky. Vacant sites occur both in the anionic and cationic sub-lattices.

In addition to the mechanism of removal of an ion from a lattice site and its penetration from cell to cell as described above, another mechanism is possible and, in fact, often takes place; this mechanism was first postulated by Frenkel'.[3] In a real crystal, and all the more so in an amorphous body, there are always, in addition to regularly arranged ions and ions which have been displaced into the lattice, vacant sites left behind by the latter ions (Fig. 3). Thermal agitation leads to the filling of various vacant sites by adjacent ions which in turn leave behind their vacant sites. This process leads to a random arrangement of ions in the lattice and their diffusion within the solid. The transition of ions in an electric field takes place predominantly in the direction corresponding to that of the acting electrical force, thus setting up an electrical current. Such a mechanism of filling vacant sites requires less energy ϵ than dissociation and transfer of ions across the lattice, and therefore it

most often predominates when the current is carried by ions in the crystal lattice proper, whereas foreign impurity ions usually move from cell to cell in the manner described earlier. However, in pure crystals, with no impurity ions, which have a loose lattice in which the atoms are not closely packed, the first mechanism of ion transfer predominates.

Schottky pointed out the possibility of formation of a neighbouring pair of vacant sites in ionic crystals.[4] A crystal lattice from which a positive and a negative ion are absent remains electrically neutral. For example, in rock salt the formation of a pair of vacant sites (Fig. 4) requires less energy than the transition of an ion into an adjacent cell. In other crystals (e.g. AgCl) the mechanism postulated by Frenkel' predominates.

Formula (8b) covers all the current mechanisms considered above.

Concentration of charge carriers

The equilibrium concentration N_i (expression (6)) of mobile charges in a crystal may also be determined from the conditions of the minimum of the appropriate thermodynamic potential. At a given temperature T and a constant volume V the equilibrium is attained when the free energy F of a unit volume is a minimum

$$F = U - TS,$$

where U is the energy and S the entropy.

We shall first consider crystals in which the charge carriers are impurity ions.

When the concentration of impurity ions is N, the concentration of ions freed by thermal agitation is N_i and the energy consumed in transferring each impurity ion into the free state is equal to ϵ, then

$$U = N_i \epsilon.$$

The entropy S may be defined as

$$S = k \log w,$$

where w is the number of all possible distributions N_i of free charges among the total number of ions N contained in unit volume

$$w = \frac{N!}{N_i!\,(N-N_i)!}.$$

At high values of N_i, $(N-N_i)$ and N, one can use Stirling's formula

$$\ln N! = N \ln N - N.$$

We have then

$$F = N_i \epsilon - kT[N \ln N - (N-N_i) \ln (N-N_i) - N_i \ln N_i].$$

The minimum of F is found from the condition

$$\frac{\partial F}{\partial N_i} = \epsilon - kT \ln \frac{N - N_i}{N_i} = 0$$

whence

$$\frac{N_i}{N - N_i} = \exp(-\epsilon/kT). \tag{9}$$

When the degree of dissociation N_i/N is small, it is possible to neglect N_i in comparison with N and express the concentration of dissociated ions by the formula

$$N_i = N \exp(-\epsilon/kT). \tag{9a}$$

We shall now consider the dissociation of ions from the main lattice. In contrast with the preceding case an increase of entropy on dissociation is not only caused by the presence of N_i ions inside crystal lattice cells, but also by the absence of an equal quantity N_i of ions which have left their lattice sites.

The number of possible distributions of N_i dissociated ions inside N_1 crystal cells is

$$w_1 = \frac{N_1!}{(N_1 - N_i)!\, N_i!}.$$

The number of distributions of N_i unfilled sites among N_2 sites nominally occupied by a given ion in an undissociated crystal is

$$w_2 = \frac{N_2!}{(N_2 - N_i)!\, N_i!}.$$

The free energy

$$F = N_i \epsilon - kT[\ln w_1 + \ln w_2].$$

The equilibrium concentration of ions is given by the condition

$$\frac{\partial F}{\partial N_i} = 0, \quad N_i = \sqrt{(N_1 N_2)} \exp(-\epsilon/2kT) \tag{9b}$$

with the same temperature dependence as in expression (6).

It should be noted that in cases when a moving impurity ion is associated with an appreciable lattice distortion, the number of such distortions should also be taken into account in the calculation of the entropy. In this case, we should obtain formula (9b) instead of (9a) for impurity ions also.

When N_i' impurity ions, with a dissociation energy ϵ' and mobility u', and N_i'' lattice ions, with an energy of dissociation ϵ'' and mobility u'', participate simultaneously in electrical conduction in a solid, the expressions for N_i and σ_i become more complex.

In the special case when the impurity ions may only become attached to defects in the crystal lattice and the lattice ions recombine only at vacant lattice sites, N_i and σ_i may be expressed by the sums

$$N_i = N_i' + N_i'', \quad \sigma_i = \sigma' + \sigma''.$$

The experimental results are usually well approximated by the formula

$$\sigma_i = \sigma_1 \exp(-E'/kT) + \sigma_2 \exp(-E''/kT) \tag{9c}$$

where the first term with a higher energy E' is ascribed to lattice ions, and the second term (with a lower energy E''), to impurity ions.

Such an interpretation of the empirical formula (9c) is not strictly accurate. Usually the forces between lattice ions are altered by the presence of impurity ions occupying a proportion of the vacant lattice sites, and, conversely, dissociation of the lattice affects the equilibrium of impurity ions.

However, the separation of expression (9c) into two terms is reasonable both at high temperatures when the dissociation of lattice ions is so pronounced that a small number of impurity ions does not have any appreciable effect, and at low temperatures when the number of free impurity ions appreciably exceeds the number of matrix ions, even at low impurity contents.

Thus, although the expression (9c) is not strictly applicable over the entire temperature range for all ratios of lattice to impurity ions, it is still true to say that at sufficiently high temperatures σ represents the electrical conductivity of the lattice proper, and at sufficiently low temperatures only the electrical conductivity due to impurities.

Temperature dependence of the conductivity

Expression (8b) may be represented readily in graphic form by taking logarithms of both sides

$$\ln \sigma_i = \ln \sigma_0 - E_0/kT$$

or in logarithms to the base ten,

$$\log \sigma_i = \log \sigma_0 - 0{\cdot}43 E_0/kT.$$

A plot of $\log \sigma_i$ against $1/T$ (Fig. 5), should give a straight line with a slope $(\tan \phi)$ related to the energy E_0

$$\tan \phi = -\frac{0{\cdot}43}{k} E_0. \tag{10}$$

Making the units along the abscissa axis $1000/T$ and expressing E_0 in eV, we obtain

$$E_0 = 0{\cdot}2 \tan \phi. \tag{10a}$$

For ionic crystals, many glasses and polymers, the energy E_0 is of the order of a few tenths of an eV. The energy of thermal vibration kT expressed in the same units has the value of 0·025–0·03 eV at room temperature. The exponent E_0/kT is therefore of the order of 10–30. At these values the multiplier $\exp(-E_0/kT)$ grows very rapidly with increase of T and depends very strongly on E_0.

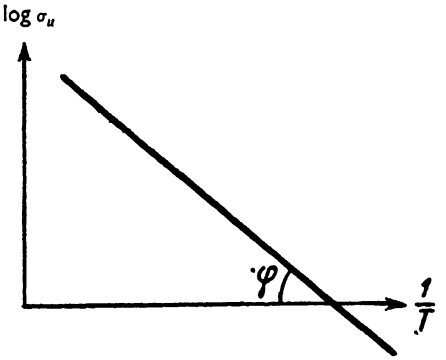


Fig. 5. Temperature dependence of the electrical conductivity of a pure ionic crystal.

Table 1 gives the values of $\exp(-E_0/kT)$ both at room temperature T_1 equal to 290° K and at twice this temperature T_2 equal to 580° K or 307° C, for different values of E_0.

TABLE 1

E_0 eV	$\exp(-E_0/kT)$	
	$T_1 = 290°$ K	$T_2 = 580°$ K
0·25	5×10^{-5}	7×10^{-3}
0·5	$2{\cdot}5 \times 10^{-9}$	5×10^{-5}
1·0	6×10^{-18}	$2{\cdot}5 \times 20^{-9}$
1·5	$5{\cdot}5 \times 10^{-26}$	$1{\cdot}2 \times 10^{-13}$
2·0	4×10^{-35}	6×10^{-18}

The coefficient $\sqrt{(N_1 N_2)}$ in expression (9b) is usually of the order of 10^{22} and σ_0 in expression (8b) is of the order of 10^5. Therefore, for E_0 equal to 1·5–2 eV, the electrical conductivity σ is too small even to be measured at room temperature, whereas at 300° C ions for which E_0 is equal to 2 eV produce currents which can be measured without difficulty.

The slope of the straight line given by formula (9a) need not correspond to E_0 at the temperature T, when the value of E_0 itself

depends on temperature. Thus, for example, when

$$E_0 = E + \alpha T$$

we have $$\log \sigma_i = \log \sigma_0 - 0{\cdot}43 \frac{\alpha}{k} - \frac{0{\cdot}43}{k} \frac{E}{T}.$$

Therefore $\tan\phi$ does not give the value of E_0 in this case but, instead, an energy E which is equal to E_0 only at the absolute zero. The relationship between $\log \sigma_i$ and $1/T$ is linear as in the previous case, but $\log \sigma_0$ is replaced by $\log \sigma_0 - 0{\cdot}43\alpha/k$.

When terms of higher order have to be taken into account in the expression for E_0, for example $E_0 = E + \alpha T + \beta T^2$, the relationship between $\log \sigma_i$ and $1/T$ represents a curve, the slope of which at temperature T cannot be taken as equal to E_0/k. In fact, it may be readily proved that

$$\frac{\partial \ln \sigma_i}{\partial (1/T)} = -\frac{E}{k} + \frac{\beta}{k} T.$$

Ion mobility

Owing to the long time required for the establishment of ionic equilibrium in crystalline quartz, Shaposhnikov and I have succeeded in separating the ion concentration N_i from the mobility u_i in the expression for the electrical conductivity σ. For this purpose we have measured the conductivity σ_i of quartz heated to temperature T, and then its conductivity σ_i' after sudden cooling to a temperature T'. It may be assumed that during such quenching there is not enough time for the number of ions to change from that existing at temperature T, whilst the mobility of the ions takes on the value corresponding to temperature T'. The conductivity of quartz kept at this temperature for a long period of time gradually decreases until its value corresponds to thermal equilibrium at temperature T'. This value of conductivity will be denoted by $\sigma_{i,0}$. It will be readily seen that the ratio σ_i/σ_i' equal to $u_{iT}/u_{iT'}$ gives the ratio of mobilities at temperatures T and T' whereas $\sigma_i'/\sigma_{i,0}$ equal to $N_{iT}/N_{iT'}$ gives the ratio of the numbers of ions at these temperatures.

Experiment yielded a somewhat higher value for the first ratio than for the second. It may be inferred from this that the change of electrical conductivity of quartz with temperature is caused both by an increase of the number of free ions, and to an even greater extent by an increase in the probability of mobile ions overcoming the barrier; thus 2ω was found to be somewhat larger than ϵ.

An attempt to determine the absolute value of the mobility in quartz yielded the value 10^{-7} cm^2/V sec. at room temperature. At 300° C the mobility rose to 10^{-4} cm^2/V sec. and at 900° C to 0·1 cm^2/V sec.

Examination of the nature of the ion carrying the current in quartz led us to the conclusion that we were dealing with impurity ions, in general, sodium ions. The mobility of Ba ions in BaO has recently been directly measured by the radioactive tracer method and found to be equal to 6×10^{-10} cm^2/V sec.

The mobility of ions depends to a large extent on their size. Thus, in quartz, in which the Si–O distance amounts to 0·82 Å, it is possible to detect the passage of sodium ions ($\bar{r} = 0{\cdot}60$ Å), whilst the mobility of the larger potassium ions ($\bar{r} = 1{\cdot}33$ Å) is too small to be measured. In sodium nitrate, I was able to observe and follow the movement of lithium ions ($\bar{r} = 0{\cdot}95$ Å), but did not succeed in driving potassium ions through the crystal.

Lukirskii has shown that in rock salt the whole current is carried at room temperature by sodium ions; the transfer of chlorine ions only becomes appreciable at temperatures exceeding 600° C (Tazulakhov's experiments[5]).

Crystalline and amorphous solids usually contain a certain amount of impurities. The dissociation energy ϵ_2 of these ions which are weakly fixed in the crystalline lattice is smaller than the dissociation energy ϵ_1 of ions which are firmly bonded in the lattice. In impurity ionic conductors, σ_i may be represented approximately as a sum of two terms, corresponding to lattice ions and impurity ions respectively,

$$\sigma_i = \sigma_1 \exp\left(-\frac{\epsilon_1 + 2\omega_1}{2kT}\right) + \sigma_2 \exp\left(-\frac{\epsilon_2 + 2\omega_2}{2kT}\right) \tag{11}$$

or, from expression (9c)

$$\sigma_i = \sigma_1 \exp\left(-\frac{E_1}{kT}\right) + \sigma_2 \exp\left(-\frac{E_2}{kT}\right).$$

Fig. 6 shows a plot of $\log \sigma_i$ against $1/T$.

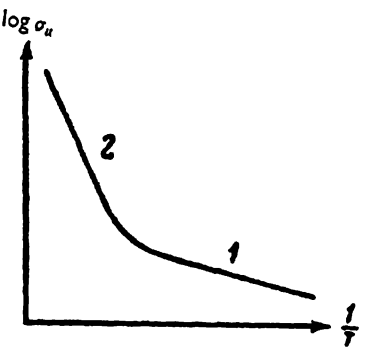


Fig. 6. Temperature dependence of the electrical conductivity of an impure ionic crystal: 1, impurity conduction range; 2, intrinsic conduction range.

Crystal lattice defects

The temperature dependence of the electrical conductivity shown in Fig. 6 has in fact often been observed. Side by side with the explanation suggested above, an alternative physical interpretation of the empirical relationship (9c) has also been advanced.

On the basis of the presence of lattice defects in real crystals, Smekal postulates that the second term in expression (9c) or (11) does not relate to chemical impurity ions but to the numerous lattice ions located on crystal defects. Quantitatively, Smekal equates the energy E_2 for attachment of the ions in each defective position with the energy required to remove an ion to the surface of the crystal, assuming E_2 is approximately a half of E_1.

The experimental value of E_1/E_2 is usually of the order of 2–3. Smekal's interpretation does not appear to me, however, to be correct. Local distortion of the lattice structures may only weaken the binding of an ion insignificantly, and in any case to a much lesser extent than a transition to the free surface where half of the adjacent ions are totally absent and where the dissociation energy may be reduced to one-half. For example, for rock salt crystals, calculations give

$$\frac{E_1}{E_2} = \frac{1{\cdot}75}{1{\cdot}63} = 1{\cdot}07.$$

In this case the energy E_2 is experimentally twice or even three times smaller than E_1.

Moreover, the energies E_1 and E_2 each consist, as we have seen earlier, of two terms, the dissociation energy ϵ and the energy barrier to displacement within the lattice, of magnitude ω,

$$E_1 = \tfrac{1}{2}(\epsilon_1 + 2\omega_1) \quad \text{and} \quad E_2 = \tfrac{1}{2}(\epsilon_2 + 2\omega_2).$$

Since according to Smekal the ions in the regions of defects do not differ in their chemical composition from the matrix ions, the barrier ω should have an equal value for the matrix ions and the defect ions (unless it is assumed that the latter move in the crystal only along the surface of internal cracks, which is exceedingly improbable). According to our experiments, i.e. for quartz, $2\omega \leqq \epsilon$.

Similar results were obtained by Koch and Wagner, who studied AgCl and AgBr with a small concentration of Cd^{++}.

Thus when E_1 is equal to $(2–3)E_2$ in the formula (9c), we have for the energy of dissociation of ions,

$$\epsilon_1 = (3–5)\,\epsilon_2.$$

It is very difficult to imagine such defects of the lattice structure which, whilst being imperceptible, would reduce the dissociation energy by 3 to 5 times, i.e. much more than open cracks.

Impurities

The postulate that the dissociation energy of impurity ions is substantially reduced is quite natural. Koch and Wagner even assumed that in their experiments ϵ could be neglected in comparison with ω. The coefficient σ_1 is usually much larger (up to 10^6 times) than σ_2, which would mean, according to Smekal, that the number of ions in the defective region amounts to approximately 10^{-6} of the total number of ions in the lattice.

In the relatively pure materials from which crystals are prepared there are always some hundredths of a per cent of impurities which are much more weakly fixed in the lattice than the matrix ions. It would be difficult to explain why these ions, the binding of which is known to be weak, should have no effect on the electrical conductivity, whereas crystal ions in the region of defects should appreciably increase the conductivity. Amongst the impurity ions there is always a certain proportion for which the transition energy ω is much smaller than for matrix ions. It is natural to expect that these ions should exert the controlling influence on the conduction in a crystal.

In fact, as we have seen earlier, the current is transferred in a quartz crystal by foreign sodium ions, and in sodium nitrate the current may be carried by lithium ions introduced into it.

Kirpicheva and I have succeeded in the preparation, by careful repeated crystallization, of alum and saltpetre specimens of so high a purity that the conductivity of all the crystals was exactly the same and could be expressed by the single term formula (8b). The electrical conductivity of the starting material, the purity of which was lower, was thousands of times higher and different for different specimens.

Our experiments show that impurity ions can carry current whenever they are present. When impurity ions are absent, then mechanical defects which are always present in real crystals do not increase the electrical conductivity, but if anything decrease it somewhat; this may be observed after the plastic deformation of a crystal which sets up a dense network of internal edge and screw dislocations (Tsekhnovitser's experiment).

Thus an analysis of the experimental material forces me to conclude that both in ionic crystals and in electronic semiconductors the extra conductivity described by the second term in expression (9c) is caused by chemical impurities and not by mechanical defects in the structure of a crystal.

When considering the problem of the different slopes of the straight lines $\log \sigma_i$ against $1/T$ at low temperatures, one more possibility should be taken into account.

In the experiments which I carried out with Shaposhnikov, we were able to freeze the ion concentration on rapid cooling of the crystal. This phenomenon appears to be widespread among solid electrolytes and semiconductors.

It may be considered that in a given crystal there exists a certain temperature range above which equilibrium is attained so rapidly that the electrical conductivity corresponds to this condition. On the other hand, below these temperatures, the time required for attaining equilibrium is so long that in an experiment one deals with a "frozen" state—namely, frozen at the concentration of free ions or vacant sites which existed at the higher temperature T_k. In this case the electrical conductivity at higher temperatures (above T_k) will depend both on the concentration and mobility of ions characteristic of the temperature at which the measurements are made, whereas at temperatures below T_k the concentration of ions and vacant sites will be independent of the temperature of measurement and will always correspond to the temperature T_k, only the mobility being a function of the specimen temperature. The slope of the straight line $\log \sigma_i$ against $1/T$ at temperatures above T_k will depend on E_1 equal to $\frac{1}{2}(\epsilon_0+2\omega)$, and below T_k on E_2 equal to ω.

The ratio of these energies is

$$\frac{E_1}{E_2} = 1+\frac{\epsilon_0}{2\omega}.$$

Some confirmation of this interpretation is given by measurements of the diffusion coefficient of ions above and below T_k carried out by the radioactive tracer method.

Chemical nature of the current carriers in ionic crystals

At elevated temperatures the currents flowing through ionic crystals are sometimes so large that it is not only possible to determine the products of electrolysis but also to measure the transfer numbers, i.e. to determine the proportions of the current carried by positive and negative ions.

As might have been expected from the pronounced dependence of current on the dissociation energy ϵ_0, it is often found that the entire current is carried by one type of ion. Thus in rock salt at temperatures below 600° C the current is carried by Na^+ cations and it is not possible to detect the participation of Cl^- ions. In contrast, in $PbCl_2$ it is the Cl^- anion that carries the current.

However, close to the melting point, the halide ions in alkali halide crystals NaF, NaCl, and KCl, also participate in current flow. Thus, according to Tazulakhov's measurements, the proportions of current in

NaCl carried by Cl^- and Na^+ ions vary with temperature in the manner shown in Table 2.

Tables 2 and 3 show that the two terms in the formula (9c) may sometimes relate to two different lattice ions and not to impurity ions. Thus, in the case of pure rock salt the temperature dependence of electrical conductivity begins to deviate appreciably from that given by the single-term formula, just at the temperatures at which the chlorine ions start to participate in the current. The measured transfer numbers are also in good agreement with this interpretation.

TABLE 2

$T°$ C	$\frac{J_{Cl^-}}{J_{Na^+}}$%
550	0± 1
600	8± 4
620	23±10
700	60±15

TABLE 3

$T°$ C	J_{Pb^+}	J_{I^-}	$\frac{J_{Pb^+}}{J_{I^-}}$%
255	0·38	0·62	60
270	0·45	0·55	82
290	0·67	0·33	200

Impurities have a double effect on the conductivity of ionic crystals. Firstly, they represent a rich source of highly mobile ions with a low dissociation energy and therefore increase by many times the electrical conductivity of such crystals; thus in solid solutions of silver chloride in sodium chloride, σ_i is 10^5 times higher than in pure sodium chloride. In this respect the impurities have the same effect on the electrical conductivity of ionic and electronic semiconductors. Secondly, impurities weaken the crystal lattice and facilitate the movement of lattice ions.

In metals the currents are carried by electrons and, therefore, the electrical conductivity of metals does not make it possible to estimate the ion mobility. However, it is possible to observe the diffusion of ions in metals (e.g. by the radioactive tracer method) and the diffusion coefficient D may then be related to the mobility by Einstein's equation (3). Thus, in the crystals of pure copper, the energy ω, which is required by an ion for its transition to an adjacent cell of the crystal lattice, is equal to 2·5 eV. Addition of 10% tin reduces this energy to approximately 1·7 eV. In lead ω drops from 1·2 to 1·04 eV on the addition of tin, and to 0·57 eV on the addition of gold. Consequently, impurities may increase the conductivity of ionic crystals even in cases when they do not participate in the current. (Let it not be thought, however, that the *electron* mobility is increased by impurities.)

In molten salts all the ions often take part in the conduction process, the ions having in this case a much higher mobility than in the solid

crystal. The fusion of salts is therefore accompanied by a steep rise of electrical conductivity which may reach the values of tens of $ohm^{-1}\,cm^{-1}$.

When comparing solid electrolytes with solutions one must not forget that, whilst the ions in solutions are surrounded during their motion by a shell of solvent, the ions in solid electrolytes represent ionized atoms. In concentrated solutions, each ion attracts an ion of the opposite sign which sets up a space charge of the opposite sign in its vicinity which weakens its own field. The effective charge of the ion, taking into account its surroundings, is smaller than the charge of the ion by itself.

Mixed ionic and electronic currents

The decisive factor for the participation in the current of a type of charge carrier is the numerical value of its energy $\epsilon_0 + 2\omega$. Thus in rock salt we have for Na^+ ions $\epsilon_0 + 2\omega \fallingdotseq 1$ eV, for Cl^- ions $\epsilon_0 + 2\omega \fallingdotseq 2$ eV and for electrons $\epsilon \fallingdotseq 6$ eV. Therefore, at room temperature, when $kT \fallingdotseq 0{\cdot}03$ eV, only Na^+ ions participate in the current. Close to the melting point at kT equal to 0·1 eV, the current is carried both by Na^+ and Cl^- ions. Electrons still do not participate significantly in the current in spite of their much higher mobility.

In other substances, such as ionic Ag_2S crystals, in which the β-Ag_2S modification changes into the α-Ag_2S modification with increase of temperature, the part played by electron conduction increases with temperature and reaches 20% at the transition point. The high-temperature α-Ag_2S modification possesses purely electronic conduction, although, as mentioned previously, for a long time it was assumed that conduction in Ag_2S was purely electrolytic. We used this example to indicate the errors which may enter into the determination of the current mechanism in solids.

In copper iodide, the low-temperature modification γ-CuI exhibits electronic conduction. Starting at 250° C, ions begin to participate in the current, and above 400° C all the current is carried by ions. γ-CuBr and CuCl behave in a similar way. In CuCl ionic conduction is appreciable even at room temperature and predominates at 300° C. In β-CuI and α-CuI, the current is virtually transferred only by ions; the proportion of electronic current is very small.

Many ionic crystals possess photoconductivity which is always due to electrons. Photoelectrons may come either from atoms forming the crystal lattice (e.g. sulphur in ZnS) or impurity atoms introduced by diffusion from outside. As a result of the action of ionizing radiation (e.g. X-rays, ultra-violet rays, cathode rays, α-rays) electrons may be localized at lattice sites from which negative ions have been removed.

In all these cases the current carried by photo-electrons is superimposed on the ionic current.

During surface oxidation of metals when metal ions diffuse across the oxide formed on the surface, the charges carried by the ions are balanced by a flow of electrons across the oxide.

In strong electrical fields, deviations from Ohm's law are observed in ionic conductors; an electronic current set up by the strong electrical field is apparently superimposed on the electrolytic current (Pruzhinina-Granovskaya's experiments[6]).

The dependence of the concentration of current on their energy of dissociation ϵ is so strong that it usually determines the nature of the current in an unequivocal manner. The case of mixed conductivity may be regarded as an exception to this law.

High-voltage polarization

Ions moving inside a crystal are trapped by impurities or localized in defective regions of the crystal, thus forming space charges. The field of these charges reduces the strength of the field set up by the external potential difference leading to a reduction of the current as the charges accumulate. When the electrodes are short-circuited without applying an external e.m.f., the field set up by the space charges causes a polarization current to flow in the reverse direction. The polarization e.m.f. is always smaller than the potential difference which has caused it, but it may reach a few thousand volts.

I have observed two types of space charges: (1) in calcite and saltpetre, the space charges accumulate at the cathode within a layer having a thickness of the order of 1 micron; (2) in quartz, space charges of both signs are arranged in thick layers close to the anode and the cathode.

Figs. 7 and 8 illustrate the potential distribution in calcite and quartz plates at various times. The lower part of Fig. 7 shows the potential distribution in calcite on short-circuiting the electrodes after the current has been allowed to flow for a long time. After 15 minutes the potential difference at the electrode which first acted as cathode dropped from an initial value of 75 V to 12 V, and a potential difference of 10 V was established at the electrode which acted first as anode and then as cathode for the polarized current. The decrease of the polarization current almost to zero does not always indicate that the polarizations at the electrodes have vanished but frequently only that they balance each other.

High-voltage polarization cannot be regarded as a symptom of an ionic current mechanism. It only indicates that there exists some mechanism which interferes with the transfer of charges to the electrodes or that there are centres, to which the charges adhere, close to the

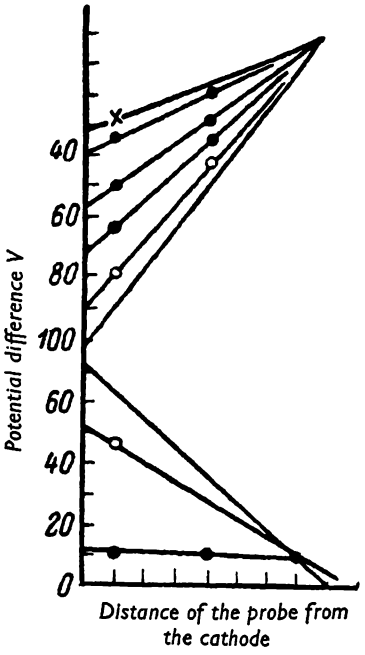


Fig. 7. Potential distribution in calcite at various times after the application of an external voltage (temperature ~150° C). The lower portion of the diagram illustrates the disappearance of polarization at zero external voltage. Owing to the reverse current a potential step is also set up at the anode.

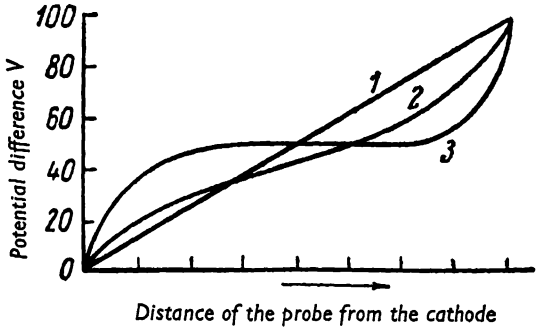


Fig. 8. Gradual formation of high-voltage polarization in crystalline quartz: 1, the initial polarization; 2, an intermediate state; 3, final potential distribution.

electrodes. For example, a photoelectric current in rock salt causes similar high-voltage polarization phenomena to those observed in the ionic conductors, quartz and calcite, although it has a purely electronic character. Similar polarization is produced by electronic currents in a diamond.

The effect of ionic currents in electronic conductors

Our main object is to describe the properties of electronic semiconductors. However, since many electronic semiconductors represent crystals formed by positive and negative ions, it is necessary to take into consideration the possibility of the presence of ionic conduction. The presence of ions in semiconductors with extrinsic conductivity is even easier to visualize since each free electron excited from an impurity atom leaves behind a positive ion.

Since the semiconductor remains electrically neutral, the number of electrons in this case is equal to the number of positive impurity ions (or the number of "holes" is equal to the number of negative ions which have picked up the electrons from the highest filled band). The mobility of electrons is 6–10 orders of magnitude higher than the mobility of ions and therefore the ionic current is practically insignificant. It may, however, affect slow processes, such as the "ageing" of rectifiers, etc.

For technical applications of semiconductors it is necessary to verify that there are no appreciable ionic currents, even if they are very small in comparison with the electrical currents, since electrolysis gradually changes the properties of the semiconductor, and at high currents may lead to its breakdown.

Many characteristics established for electrolytes are equally valid as far as the electronic properties of semiconductors are concerned. The physical sense of these laws is easiest to understand in terms of the motion of ions, since, in the case of electrons, it is their quantum-mechanical properties that are of decisive importance. These properties manifest themselves with particular clarity, when comparing similar phenomena in electronic conductors and solid electrolytes.

References

GENERAL

A. F. Ioffe, *The Physics of Crystals*. McGraw-Hill Book Co., New York, 1928.

N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals*. Clarendon Press, Oxford, 1950.

A. B. Lidiard, "Ionic Conduction", *Handbuch d. Physik*, Vol. XX, *Electrical Conductivity*—II. Springer-Verlag, Berlin, 1957.

SPECIFIC PROBLEMS

1. F. Klaiber, "Hall effect and conductivity of silver sulphide", *Ann. d. Phys.*, 3, 229 (1929).
2. C. Wagner, "Nature of electrical conductivity in α-silver sulphide", *Zs. f. Phys.* (B), 23, 469 (1933).
3. Ya. I. Frenkel', "Thermal motion in solids and fluids", *Zs. f. Phys.*, 35, 652 (1926).

4. W. Schottky, "The mechanism of the motion of ions in solid electrolytes", *Zs. f. phys. Chem.* (B), **29**, 335 (1935).
5. P. I. Lukirskii, "The conductivity of natural and ionised rock salt", *Zhur. Rus. Fiz. Khim. Obshch.*, ch. fiz., **50**, 26 (1919).
6. V. I. Pruzhinina-Granovskaya, "The nature of electrical conductivity of mica in strong electric fields", *Zhur. Eksper, Teor. Fiz.*, **7**, 506 (1937).

CHAPTER

2

Metals

The motion of electrons

The passage of current through metals is not accompanied by chemical changes. In this case the current carriers always consist of electrons which have a charge e of $-1{\cdot}6 \times 10^{-19}$ coulombs. As in other conductors, the specific electrical conductivity σ of a metal may be expressed by formula (1) through its mobility u and the concentration n of electrons participating in the current

$$\sigma = neu.$$

We regard the mobility as the average velocity of all the free electrons in the direction of a field of unit strength. We shall see later that not all the electrons in the metal, by any means, contribute equally to the current. In the simplified semi-classical treatment to be adopted in this chapter we shall, however, regard the current as a displacement of the totality of the electrons with a certain average velocity $u\mathscr{E}$. We regard the description of the properties of metals given in this chapter as semi-classical since we use the quantum theory of metals only to obtain the statistics. The electrons themselves and their interaction with the crystal lattice will be regarded as classical without taking into account the quantum nature of the movement of electrons in the periodic lattice field.

When n and u are independent of the electrical field strength, as in the case of metals, formula (1) represents Ohm's law. This law remains valid for metals up to very high current densities.

Since neither the number of electrons n nor their mobility u are directly known to us, we have to use another measurable quantity related to n and u for their separation. One can use the Hall coefficient R for this purpose.

With a magnetic field H (Fig. 9) perpendicular to the direction of the current J, an electrical field $\mathscr{E}_H$, normal to both H and J, is set up which is given by the expression

$$\mathscr{E}_H = RJH. \tag{12}$$

The direction of the field $\mathscr{E}_H$ depends on the sign of the charges carrying the current. Since all electrons have a negative charge, the direction of $\mathscr{E}_H$ might be expected to be the same for all metals. However, while experiment gives the predicted direction of the field $\mathscr{E}_H$ for some metals, for others $\mathscr{E}_H$ is in the opposite direction. The physical meaning of this will be considered later in conjunction with similar phenomena in semiconductors.

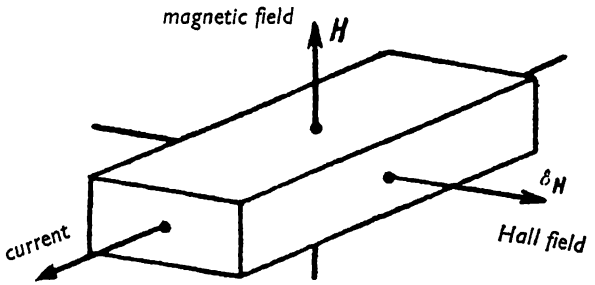


Fig. 9. Transverse field $\mathscr{E}_H$ due to the Hall effect.

From elementary considerations of electronic theory, the constant R is given by

$$R = \frac{1}{nec}. \tag{13}$$

Electron mobility

Multiplying the left-hand and the right-hand side of equations (1) and (13), we obtain

$$\sigma R = \frac{u}{c}. \tag{14}$$

Experiments show that the product σRc on which electron mobility depends is of the same order of magnitude (a few tens or hundreds of cm/sec in a field strength of 1 V/cm) for metals and for many electronic semiconductors.

The electrical conductivity σ of metals ranges from 10^5 to 10^6 ohm^{-1} cm^{-1}, and the electronic charge e is $-1{\cdot}6 \times 10^{-19}$ coulomb. The number of electrons carrying the current is found to be equal to about 10^{22} to 10^{23} per cm^3. This value is of the same order as the number of atoms in 1 cm^3. On the assumption that all the valence electrons of the metallic atoms participate in the current, we obtain an average mobility of

$$\bar{u} \doteq (10\text{–}1000)\ \text{cm}^2/\text{V sec}.$$

We consider the model as a totality of positive ions distributed at crystal lattice sites within the metal, at certain distances from each

other. The remaining space in the metal is occupied by electrons which may pass from one ion to another without any resistance. The potential distribution in this volume is extremely non-uniform; in the portions occupied by positive ions the potential energy of an electron drops abruptly, whilst in the spaces between the ions it increases. The width a of these interatomic potential barriers is approximately equal to 3×10^{-8} cm. At a height of 1 eV for these interatomic barriers and for the appropriate wave function of an electron covering such a distance, the barriers cannot have any effect on the movement of the electrons. According to quantum theory, even when the distances a are large, electrons in the strictly periodic field of a crystalline lattice do not meet any resistance to their translational motion.

Such conditions would exist in the vicinity of the absolute zero in perfect crystals of an absolutely pure metal. The electrical conductivity becomes high at low temperatures because the electron mobility rises.

The effect of temperature on electron mobility

Thermal agitation affects the regularity of the lattice and causes the ions to approach each other and to move apart, therefore producing fluctuations of the electrical field extending over a few atomic distances.

Electron waves are scattered by these fluctuations; such scattering corresponds to a change of the momentum and sometimes, also, of the energy of electrons regarded as discrete particles.

The thermal motion of ions may be visualized as a collection of sound waves. A scattering act may result not only in a change of direction of motion of the electron, but also in a transfer of energy $h\nu_s$ from or to the crystal lattice where ν_s is the frequency of one of the standing sound waves. The largest number of frequencies of thermal vibration are close to the limiting frequency $\nu_{m,s}$ with a wavelength λ of about 5×10^{-8} cm. For a velocity of sound of the order of 5×10^5 cm/sec, this gives for $\nu_{m,s}$ a value of approximately 10^{13} c.p.s. and for $h\nu_{m,s}$ approximately 4×10^{-14} erg, or 3×10^{-2} eV, this corresponding to the average energy of thermal vibration at room temperature. For this reason, electrons lose or acquire energy, during scattering by thermal fluctuations of the lattice, most often in amounts of the order of 3×10^{-2} eV. The density and intensity of the fluctuations and, at the same time, the probability of a change of momentum and transfer of energy by the electrons, increase with increasing thermal vibration. The velocity and kinetic energy of an electron which has moved freely for some time in an electrical field change under the action of the electrical forces. One can simplify the real process of scattering by assuming that an electron will travel a distance l between the scattering acts, when it changes its direction of motion and sometimes also its

energy (the latter by $h\nu_s$, amplifying or attenuating the thermal oscillations of the crystal lattice). The smaller the free path length l, the shorter is the time τ during which the field accelerates the electron, the smaller its average additional velocity Δv in the direction of the electrical field, which is superimposed on its random motion, and the smaller the mobility u. The latter depends only on the average value of the additional velocity Δv acquired by the electron in the electrical field, Δv being very small in comparison with the thermal velocities.

The velocity acquired by a free electron moving in an electrical field $\mathscr{E}$ under the effect of a force $e\mathscr{E}$ will amount at the end of the period τ to

$$\Delta v_{\max} = \frac{e}{m}\mathscr{E}\tau$$

and the average velocity increment in the direction of the field will be

$$\overline{\Delta v} = \frac{1}{2}\frac{e}{m}\mathscr{E}\tau.$$

The mobility u, i.e. the average velocity of translation for all the electrons in a unit field is†

$$u = \frac{1}{2}\frac{e}{m}\tau \tag{15}$$

and the electrical conductivity for a concentration of electrons n equal to N_0 is

$$\sigma = \tfrac{1}{2}N_0\frac{e^2}{m}\tau.$$

The electrons in the metal move in a medium with a periodically varying potential and therefore it is not at all obvious that their motion obeys the same laws of mechanics as motion in a free space. Quantum mechanical considerations lead, however, to the conclusion that the average motion of electrons in a crystal obeys these laws provided the mass of an electron m_0 is replaced by a certain quantity m^* which has been given the name effective mass. For the majority of metals m^* does not differ much from m_0.

Thermal conductivity

Electrons in a metal not only carry the electrical current but also contribute to the heat flow, increasing the thermal conductivity above that due to the lattice. This phenomenon was interpreted by the classical theory of metals in the following manner.

† Taking into account the statistical nature of τ, one gets the correct formula

$$u = \frac{e}{m}\bar{\tau}. \tag{15a}$$

Electrons in equilibrium with the hottest parts of the metal possess the highest average energy. Passing into cooler regions the electrons transport this additional energy and raise the temperature of these parts. We shall consider the case when the temperature T varies in the direction x but is constant in the plane yz.

The amount of heat Q transferred from the hotter portion of the crystal to the cooler portion is proportional to the cross-section area S of the thermal flux and the temperature gradient $\partial T/\partial x$

$$Q = \kappa S \frac{\partial T}{\partial x}. \tag{16}$$

The coefficient κ is known as the thermal conductivity. The kinetic theory of gases gives

$$\kappa = \tfrac{1}{3} l \bar{v} c'_v \tag{17}$$

where l is the free path length of gaseous molecules, $\bar{v}$ their average velocity of motion, and c'_v the thermal capacity per unit volume of the gas at constant volume.

By an analogy with the case of a gas we shall apply expression (17) to electrons in a metal. The quantities entering the expression have the following values:

$$l = \bar{v}\tau.$$

The average velocity $\bar{v}$ of movement of the electrons will be assumed to correspond to the average kinetic energy of gaseous molecules (following the erroneous classical electronic theory)

$$\tfrac{1}{2} m \bar{v}^2 = \tfrac{3}{2} kT \tag{17a}$$

and the thermal capacity

$$c'_v = \tfrac{3}{2} N_0 k$$

where N_0 is the density of the electrons transporting the heat energy. We then have

$$\kappa = \frac{3}{2} \frac{N_0 \tau k^2 T}{m}. \tag{18}$$

At the same time the electrical conductivity σ may be expressed with the aid of equation (15) as

$$\sigma = \tfrac{1}{2} N_0 \frac{e^2}{m} \tau. \tag{19}$$

Hence the ratio of thermal conductivity κ to electrical conductivity σ should be

$$\frac{\kappa}{\sigma} = 3 \frac{k^2}{e^2} T. \tag{20}$$

If we assume that the velocity distribution of electrons is the same as that of molecules in a gas, we get

$$\frac{\kappa}{\sigma} = 2\frac{k^2}{e^2}T. \tag{20a}$$

According to equation (20a), the ratio κ/σ is independent of both N_0 and τ and, therefore, should be equal for all metals at a given temperature. At 300° K (assuming k is $1{\cdot}38 \times 10^{-16}$ erg/deg, or $1{\cdot}38 \times 10^{-23}$ J/deg, and e is $-1{\cdot}6 \times 10^{-19}$ coulombs) this ratio has the value

$$\kappa/\sigma = 4{\cdot}5 \times 10^{-6}$$

where κ is expressed in W/deg cm and σ in ohm^{-1} cm^{-1}.

In fact, Wiedemann and Franz, who measured the thermal and electrical conductivity of a number of metals, reached the conclusion that the ratio of these quantities is approximately the same for all metals and the actual value of this ratio was found to be quite near to that calculated above.

As we shall see later, the main assumption in the foregoing calculation, according to which the energy of the electrons is equal to the energy of gaseous molecules at the same temperature, is not true when applied to metals but is quite applicable for electrons in semiconductors. The statement that all the N_0 electrons in the metal participate in the current flow is also incorrect. Incidentally, this assumption does not affect the ratio of κ/σ since N_0 cancels out. The result given by expression (20a) is changed but little by the quantum theory taking the form

$$\frac{\kappa}{\sigma} = \frac{\pi^2}{3}\frac{k^2}{e^2}T. \tag{20b}$$

The ratio κ/σ in a metal is greater than that calculated earlier by $2\pi^2/3$ or 1·64 times, whereas expression (20a) remains true for semiconductors where the velocity distribution of electrons is the same as for gaseous molecules, obeying the Maxwell–Boltzmann law. Actually, the coefficient of $(k^2/e^2)\,T$ in the expression for κ/σ is governed by the way in which the scattering of the electrons depends on their velocity and takes on values ranging from 1 to 4.

Expression (20a), as well as expression (20b), relates only to that portion of thermal conductivity which is due to the movement of electrons. Both expressions take no account of the lattice thermal conductivity. The latter plays the dominant role in semiconductors, but is unimportant for the majority of metals, with the exception of such light and hard metals as beryllium. When the scattering of thermal waves by electrons may be neglected (at electron concentrations of less than 10^{20}/cm^3) the thermal conductivity κ of an electronic semiconductor

may be expressed as the sum of the thermal conductivity due to the sound waves in the crystal lattice κ_1, and the thermal conductivity due to electrons κ_2, which is given for metals by expression (20b) or for semiconductors by the corrected formula (20a).

It is important to remember, however, that, in fact, the presence in the metal of a large number of free electrons in turn reduces the value of κ_1 owing to the additional scattering of heat waves on electrons. Therefore the separation of the thermal conductivity of a metal into two terms κ_1 and κ_2 represents quite a complex problem which has not yet been solved.

Temperature dependence of the conductivity

The free path length of an electron in a metal is of the order of 10^{-6} cm. The field strength for a current density of 10 A/cm^2 and an electrical conductivity σ of 10^5 ohm^{-1} cm^{-1} is 10^{-4} V/cm. The energy accumulated along a free path length is equal to 10^{-10} eV, whereas the average kinetic energy of electrons in a metal is equal to a few eV. Therefore, the small change of energy in covering a single free path has virtually no effect on the energy of electrons in the metal. Electrons interacting with the lattice impart to it, or acquire from it, energy in amounts $h\nu$ of the order of hundredths of an eV.

The higher the temperature the more violent is the thermal agitation and the more frequent are the fluctuations of the electrical field of the atoms in the lattice which produce appreciable scattering of electronic waves and, therefore, the lower is the mobility. Since the number of conduction electrons in a metal does not vary with temperature, it follows from expression (1) that the temperature dependence of σ is the same as that of u.

It is shown by calculation and confirmed by experiment that the mobility u of electrons and the electrical conductivity σ in a solid metallic crystal are inversely proportional to its thermal energy. This statement is true for temperatures exceeding the Debye temperature θ, although it is also found to hold at somewhat lower temperatures.

Let us denote the specific heat of the metal by c. We can then assume that

$$\sigma = A \Big/ \int_0^T c\,dT,$$

where A is a coefficient of proportionality.

This relationship is found to agree well with experiment in the case of perfectly pure specimens of the metals Al, Cu, etc., in which the thermal agitation of the crystal lattice is the only source of electron scattering; an exception is found at very low temperatures in which we

have for the thermal oscillations of the lattice,

$$\int_0^T c dT \propto T^4,$$

and σ is inversely proportional to T^5.

At temperatures substantially higher than the Debye temperature θ the thermal capacity c becomes a constant equal to c_0, and the resistance of the metal almost proportional to the absolute temperature. The resistivity ρ may in this case be expressed in the form

$$\rho = \frac{1}{\sigma} = \frac{1}{A}\left[\int_0^\theta c dT + c_0(T-\theta)\right] = \frac{1}{A}\left[c_0 T - \int_0^\theta (c_0 - c)\, dT\right];$$

the value of $A\rho$ is the closer to $c_0 T$ the higher the temperature T and the larger $c_0 T$ as compared with the constant

$$\int_0^\theta (c_0 - c)\, dT.$$

The scattering of electronic waves may occur not only on pronounced non-homogeneities of the electrical field caused by the fluctuation of the thermal motion of atoms in the lattice, but also on the even more pronounced distortions of the field produced by extraneous impurity atoms. When the number of such atoms is large, as in the case of alloys, representing solid solutions, the electron mobility and electrical conductivity decrease. Metal alloys as a rule possess a higher resistivity with a lower temperature coefficient than for a pure metal.

It is necessary to distinguish between two fundamentally different arrangements of atoms of one metal among atoms of another metal which together form the alloy: (1) an ordered arrangement when the two types of atom alternate in a strictly defined way forming periodically arranged layers, such as, for example, are formed by Na^+ and Cl^- ions in a rock salt crystal; (2) a random arrangement of the two types of atom which prevails in an alloy at temperatures above a certain critical temperature.

An ordered alloy may be regarded as a regular crystal lattice, consisting of superimposed structures. Such a lattice exhibits periodicity and therefore does not present a resistance to the free movement of electrons and does not create additional scattering centres. The resistance of an alloy in the ordered state and its entropy are smaller than in the disordered state.

In ferromagnetic metals another type of ordered structure is formed, consisting of domains which contain groups of electrons with the same

orientations for mutually related spins and magnetic moments. In antiferromagnetic materials domains bring together groups of electrons with pairs of opposite orientation. On reaching a certain critical temperature (the Curie point), the bonds within the domains are broken and the distribution of spins becomes random. Both in the case of ferromagnetic and antiferromagnetic materials this leads to a transition into the paramagnetic state.

The effect of impurities on the mobility and, therefore, also on the electrical conductivity, becomes particularly pronounced at low temperatures at which the scattering by fluctuations of the thermal motion becomes smaller. Close to the absolute zero a minute quantity of impurities may bring about a marked change in the conductivity. In such metals, which are not entirely pure, the scattering centres at sufficiently low temperatures consist mainly of impurity atoms. Their number does not change with temperature and, therefore, the electron mobility and electrical conductivity also remain virtually constant at these temperatures.

In order that the scattering of electrons shall result from the thermal vibrations of the lattice down to the very lowest temperatures, the metal must be of exceptional purity. The electrical conductivity in such metals reaches very high values in the vicinity of the absolute zero, being proportional to the fifth power of the absolute temperature.

Some metals such as mercury and lead, and also some compounds such as some carbides and nitrides which possess much higher resistivities than metals, undergo a transition into the superconducting state on cooling.

A characteristic feature of metals is not only the higher concentration of conduction electrons, but also the independence of their number and the weak dependence of their energy on temperature.

Electron theory of metals

The first electron theory of metals, proposed by Rieke and Drude, and later extensively developed by Lorentz and Debye, postulated that electrons in a metal move freely, in the same way as gas molecules and, like the latter, equate their energy with the thermal energy of the atoms of the metal in which they are contained.

According to this theory the mean kinetic energy of electrons, like the energy of gas molecules, was considered to be equal to $\frac{3}{2}kT$, where T is the absolute temperature and k is the Boltzmann constant.

This theory explained Ohm's law satisfactorily, and also the relationship between the electrical and thermal conductivities of metals (Wiedemann–Franz law) and the rise of resistivity ρ with temperature, although according to a law ($\rho \propto \sqrt{T}$), different from that given by

experiment ($\rho \propto T$). It seemed that the measured velocities of electron emission from a metal represented a direct proof of the validity of the main hypothesis which compared the electrons in a metal with the molecules of a gas, since the resulting velocities were found to coincide with the velocities of gas molecules.

The electron theory led, however, to an unavoidable disagreement with the experimental values for the specific heats of metals; the law of the uniform distribution of energy among the degrees of freedom made one expect that in addition to the three degrees of freedom for each atom in the metal there would be three degrees of freedom for each free electron. As we have seen earlier, the number of free electrons had to be regarded as equal to the number of atoms. By attributing to each degree of freedom a sum of potential and kinetic energy kT, and therefore a thermal capacity k, one would expect that a gram-atom of a substance containing N_0 (equal to $6{\cdot}02 \times 10^{23}$) atoms would have a thermal capacity of $3kN_0$ or about 6 cal in the absence of free electrons in the substance, and approximately 9 cal for metals. Experiment shows, however, that the thermal capacity of a gram-atom of a metal is the same as that of an insulator.

Similar contradictions also arose in the theory of specific heats for all substances at low temperatures, where the atomic specific heat drops well below 6 cal and at very low temperatures approaches zero.

The law of uniform energy distribution among the degrees of freedom also led to a strong disagreement with experiment in the theory of radiant energy. The number of possible natural oscillations within a given frequency range is proportional to the square of the frequencies and, therefore, the energy for a given frequency interval contained in radiation in thermal equilibrium with a body, should be proportional to the square of the frequency. Experiment shows, however, that this is true only at low frequencies; at higher frequencies the energy density reaches a maximum and then diminishes, approaching zero.

All these contradictions between theory and experiment were eliminated by the quantum theory.

In 1927 there appeared two theories of metals based on the principles of quantum mechanics: (1) Frenkel's theory in which the electrons in the metal were considered to be at certain definite quantum levels but could freely exchange their places; and (2) Sommerfeld's theory, which, like Drude–Lorentz theory, started with the concept of an electron gas in the metal, but made it obey different laws of quantum statistics.

Sommerfeld's theory, which will be outlined in this chapter, neglects the periodic variation of the electrical potential in the lattice, regarding the potential as constant in the whole metal.

Quantum statistics of electrons in metals

Boltzmann's classical statistics with a uniform distribution of energy among the degrees of freedom was replaced by the quantum statistics, which Fermi and Dirac evolved for the electrons in a metal.

According to quantum statistics: (1) individual electrons in a system are indistinguishable; (2) changes of their state are determined by a change of the quantum number by integral steps; (3) in addition, it is maintained that in any complex electron system there can be no more than one electron with a given direction of spin in every quantum state; (4) there are only two possible spins, these having the quantum numbers of $\frac{1}{2}$ and $-\frac{1}{2}$. Starting with these assumptions we shall derive the law of energy distribution among electrons in a metal.

The state of the electron will be described by its three coordinates q_1, q_2 and q_3, and the three projections of the momentum p_1, p_2, p_3, along the three coordinate directions. In momentum space we shall isolate a surface (Fig. 10) corresponding to all states with the given energy E. This will be the surface of a sphere with a certain radius P, determined by the familiar relationship

$$E = \frac{P^2}{2m} \tag{21}$$

where m is the electron mass and $P^2 = p_1^2 + p_2^2 + p_3^2 = 2mE$.

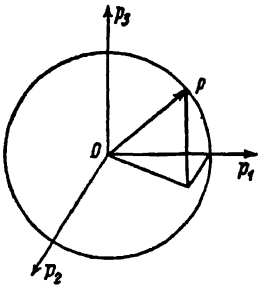


Fig. 10. Momentum space.

We have then

$$dP = \tfrac{1}{2}\sqrt{(2m)}\, E^{-\frac{1}{2}}\, dE. \tag{22}$$

The whole space of momenta and coordinates will be divided into cells with a volume h^3. According to the laws of quantum statistics listed above, the number of cells will characterize the number of possible different states. We shall carry out integration with respect to q_1, q_2 and q_3, and isolate in a unit volume all states within the range from E to $E+dE$. In momentum space (Fig. 11) these states will correspond to a spherical shell with a radius P and a thickness dP; its volume is $4\pi P^2\, dP$.

The number of cells z in this layer is

$$z = \frac{4\pi P^2\,dP}{h^3} = \frac{4\pi \times 2mE\sqrt{(2m)}\,E^{-\frac{1}{2}}\,dE}{2h^3} = \frac{2\pi\,(2m)^{\frac{3}{2}}\,E^{\frac{1}{2}}\,dE}{h^3}. \qquad (23)$$

According to the rules of Fermi statistics outlined at the beginning of this paragraph, there can be one electron, or two electrons with opposite spin directions, in each of a certain number of cells n out of the total number z; the remaining $(z-n)$ cells do not contain electrons. The number of cells in the space p, q, corresponding to a volume V of a metal will be V times larger than in the space corresponding to unit volume, but the volume of a cell in momentum space will be V times smaller:

$$\iiint dq_1\,dq_2\,dq_3 = V; \quad \iiint dp_1\,dp_2\,dp_3 = \frac{h^3}{V}.$$

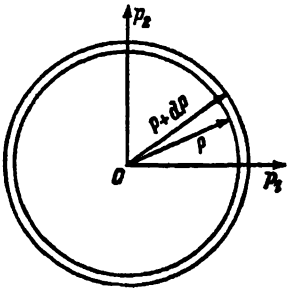


Fig. 11. A spherical shell in momentum space.

Velocity distribution of electrons

The dependence (23) of the number of quantum cells z on the energy of the electrons is illustrated in Fig. 12, in which the abscissa axis has been divided into equal energy intervals ΔE, whilst the ordinate axis shows the corresponding number of cells z.

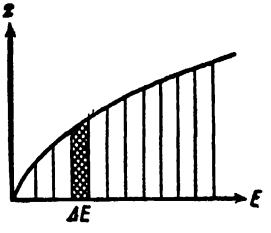


Fig. 12. Distribution of quantum cells with energy.

We shall consider first the electron distribution in a metal among the cells of phase space at the absolute zero of temperature. Here the energy

corresponds to the free energy, the minimum of which determines the equilibrium of the n_0 electrons contained in unit volume of the metal. These electrons, tending to pass into the lowest energy states, will occupy the $n_0/2$ cells with the lowest energy.

The distribution of all the n_0 electrons among the cells will be expressed by the diagram in Fig. 13 in which the number of electrons in a cell is plotted against the number of cells arranged according to increasing energy; the number of electrons in all the occupied cells is equal to 2, but starting with cell $[(n_0/2)+1]$ it is equal to zero. Therefore, in all the cells occupied by the electrons, the number of electrons

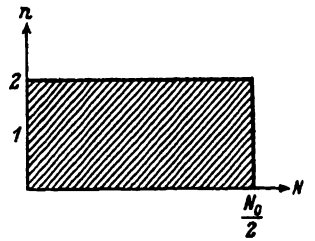


Fig. 13. Distribution of electrons among the cells of phase space, $T = 0°$ K.

with an energy in the range from E to $E+dE$ is equal to double the number of cells with this energy. According to expression (23) and Fig. 12, the number z increases with the electron energy. We shall calculate the maximum energy E_{max} of electrons corresponding to the phase cell $n_0/2$. For this purpose we shall add up (and since the dimensions of the cells are very small the sum may be replaced by an integral) the numbers n equal to $2z$ for all the shells in momentum space with energies in the range from $E = 0$ to $E = E_{max}$, and equate their sum to the total number of electrons n_0 in unit volume:

$$\frac{4\pi(2m)^{\frac{3}{2}}}{h^3}\int_0^{E_{max}} E^{\frac{1}{2}}\,dE = n_0. \tag{24}$$

Integration gives

$$\frac{4\pi(2m)^{\frac{3}{2}}}{h^3}\,\frac{2}{3}E_{max}^{\frac{3}{2}} = n_0 \tag{24a}$$

where

$$E_{max} = \frac{h^2}{2m}\left(\frac{3}{8\pi}n_0\right)^{\frac{2}{3}}. \tag{25}$$

Substituting the values h equal to $6{\cdot}6 \times 10^{-27}$, m equal to 9×10^{-28} gm and n_0 equal to 6×10^{22} (which approximately corresponds to the number of electrons in 1 cm^3 of silver) in expression (25) we obtain

$$E_{max} = 14{\cdot}4 \times 10^{-12}\ \text{erg} \doteq 9\ \text{eV}.$$

The mean electron energy $\bar{E}$ at the absolute zero may be found by dividing the energy of all electrons by their number n_0:

$$\bar{E} = \frac{4\pi(2m)^{\frac{3}{2}}}{n_0 h^3} \int_0^{E_{\max}} E^{\frac{1}{2}}\, dE = \frac{3}{5}\frac{h^2}{2m}\left(\frac{3n_0}{8\pi}\right)^{\frac{2}{3}} = \frac{3}{5} E_{\max}. \tag{26}$$

Using the same assumption as above we find $\bar{E}$ equals about 5·4 eV.

If all the electrons in the metal obeyed classical statistics and their mean energy were $\frac{3}{2}kT$, we should find at the absolute zero that $\bar{E}$ would be zero.

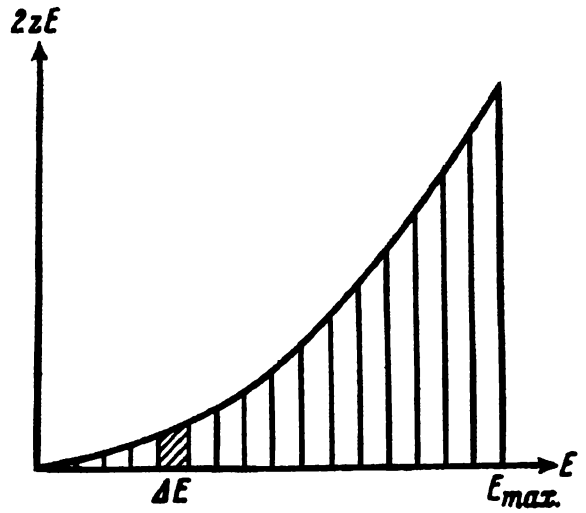


Fig. 14. The distribution of electron energies at $T = 0°$ K.

We shall plot the energy distribution of electrons at zero temperature among the various energy states. As in Fig. 12, we shall mark equal energy intervals ΔE along the abscissa axis and, along the ordinate axis, instead of the number of cells z, the energy $2zE$ of the electrons contained in the cells (Fig. 14). Since the number of levels $n_0/2$ is extremely large (of the order of 10^{22}) and all the levels are distributed between zero and 5–10 eV according to the law (23), the average distance between adjacent levels will be approximately 10^{-22} eV. We shall not be much in error if we will replace all the discrete levels by a continuous band filled by electrons with a density increasing as E increases from zero to $E_{\max}$. Fig. 15a shows the energy distribution of electrons at 0° K.

The effect of temperature

We shall now consider the changes in the energy distribution, among the electrons in a metal, brought about by a rise of temperature.

At a certain temperature $T>0$ some of the electrons, after picking up additional thermal energy, may pass into cells with higher energy, vacating an equal number of cells with lower energy. Since the average energy imparted to electrons by the thermal vibrations at room temperature amounts only to about 0·03 eV, and all the levels up to an energy E_{max} of about 5–10 eV are occupied, most of the electrons capable of passing into higher levels should possess energies differing from E_{max} only by a quantity of the order of hundredths of an eV. At the same time the surplus energy received by the electrons as a result of thermal agitation will also be limited to hundredths of an eV.

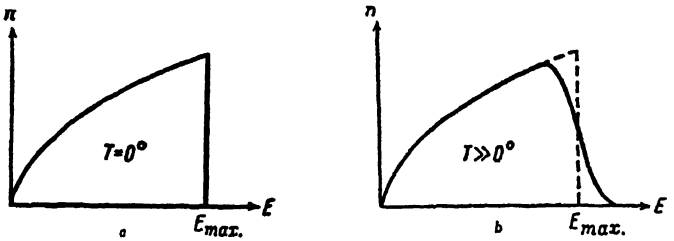


Fig. 15. Distribution of electrons with energy (a) at $T=0°$ K, (b) at $T\gg 0°$ K.

It may thus be expected that the effect of thermal agitation on the distribution of electrons in a metal will be limited to a change in the energy of electrons close to the upper limit E_{max}, known as the Fermi level. A fraction of the energy levels a few hundredths of an eV below this level will be vacated, and the same fraction of the corresponding levels above E_{max} will be occupied (Fig. 15b). The further we move, in comparison with kT, away from the Fermi level in one or the other direction the smaller will be the difference between the electron concentration at the temperature T and the concentration at the absolute zero. This circumstance explains the low specific heat of electrons in metals: the heat energy is imparted only to a small fraction of the electrons, namely those electrons with an energy lying within kT of the Fermi level. At room temperature the proportion of electrons is about 1% of the total number.

Energy distribution of electrons

To find out the most probable energy distribution of electrons at the temperature T we shall calculate the number of possible distributions permitted by Fermi–Dirac statistics and establish the conditions under which the total free energy is a minimum. These conditions lead to the equilibrium distribution. Unlike in classical statistics, in quantum statistics we consider the volume of the cells to be finite and equal

to h^3; the states are only distinguished by the number of electrons in a given cell, and no account is taken of the change of places by the electrons, since the replacement of one electron by another electron does not bring about any changes in the physical state of the metal.

The number w_i of different arrangements of a given number of electrons n_i, with a certain energy E_i, for one direction of spin, in z_i quantum cells under the condition that no single cell should contain more than one electron with a given direction of spin, is equal to the number of combinations of n_i elements from a total of z_i elements

$$w_i = \frac{z_i!}{(z_i - n_i)!\, n_i!}.$$

The number of arrangements W among all the cells is equal to the product of the numbers w_i corresponding to each individual case

$$W = \prod_i \frac{z_i!}{(z_i - n_i)!\, n_i!}. \tag{27}$$

The free energy F will be expressed in the form

$$F = U - TS = \sum_i (n_i E_i - kT \ln w_i). \tag{28}$$

To find out the equilibrium distribution of electrons it is necessary to equate the derivative of F with respect to n_i to zero. In doing this, it must be remembered, however, that the total number of electrons must remain unchanged. Therefore we shall add to F the term $-\mu n_i$ containing an arbitrary factor μ and equate the derivative of $(F - \mu n_i)$ to zero

$$\delta \sum_i \left(n_i E_i - \mu n_i - kT \ln \frac{z_i!}{(z_i - n_i)!\, n_i!} \right) = 0. \tag{29}$$

According to Stirling's formula, at high values of z

$$z! \fallingdotseq z^z \exp(-z).$$

This gives $$E_i - \mu = kT \ln \frac{z_i - n_i}{n_i}$$

and omitting the subscript i in E_i

$$n_i = \frac{z_i}{\exp\left(\frac{E - \mu}{kT}\right) + 1}. \tag{30}$$

The number of electrons per quantum state, or rather the probability p that the state will be filled, is obtained by equating z to unity.

This gives

$$p = \frac{1}{\exp\left(\frac{E-\mu}{kT}\right)+1}. \tag{30a}$$

The value of μ may be determined from the condition

$$\sum_i n_i = n_0$$

which gives

$$\mu = \left(\frac{\partial F}{\partial n_0}\right)_{V.T.}.$$

Inserting the value of z_i from expression (23) into expression (30) we obtain a law of energy distribution of electrons

$$dn = \frac{4\pi(2m^*)^{\frac{3}{2}}}{h^3} \cdot \frac{E^{\frac{1}{2}}}{\exp\left(\frac{E-\mu}{kT}\right)+1} dE, \tag{31}$$

where m^* is regarded as the effective mass of an electron in its motion in the periodic lattice field.

Replacing $-(\mu/kT)$ by λ, we have

$$dn = \frac{4\pi(2m^*)^{\frac{3}{2}}}{h^3} \cdot \frac{E^{\frac{1}{2}}}{\exp\lambda \exp\frac{E}{kT}+1} dE. \tag{31a}$$

At low values of E where $E < \mu$, and in addition $(\mu - E) \gg kT$, expression (30a) gives $p \doteq 1$. For very high values of E at which $(E-\mu) \gg kT$, we obtain $p \doteq 0$. When $E = \mu$, $p = \frac{1}{2}$. Finally, when $(E-\mu)$ is of the order of magnitude of kT, the probability of the quantum state being filled is within the range from zero to unity. The higher the value of kT, the wider is the energy interval in which $(E-\mu)$ satisfies the last condition.

In practice the distribution may be regarded as classical as long as $(E-\mu) > 2kT$, and degenerate when $(E-\mu) < -2kT$.

Therefore, the distribution of electrons at a temperature T will have the shape shown in Fig. 15b, which differs somewhat from that corresponding to the absolute zero. Not all the levels below the Fermi level E_{max} will be filled with electrons; in the vicinity of the Fermi level, at a distance of the order of kT below it, vacant levels will appear, whilst some of the electrons will occupy levels above E_{max} to approximately the same distance kT from the Fermi level.

The higher the temperature the wider is the energy transition range between a probability of unity and a probability of zero for a state to be filled.

Degeneracy of the electron gas

When $\exp\lambda$ is so large that $\exp\lambda\exp(E/kT) \gg 1$, expression (31a) leads to the Maxwellian distribution of velocities of electrons, identical with the velocity distribution of gas molecules.

In fact, to determine $\exp\lambda$ we have in this case

$$\frac{4\pi(2m^*)^{\frac{3}{2}}}{h^3}\exp(-\lambda)\int_0^\infty E^{\frac{1}{2}}\exp(-E/kT)\,dE = n_0. \tag{32}$$

But
$$\int_0^\infty E^{\frac{1}{2}}\exp(-E/kT)\,dE = \frac{\sqrt{\pi}}{2}(kT)^{\frac{3}{2}},$$

whence
$$\exp(\mu/kT) = \exp(-\lambda) = \frac{n_0 h^3}{(2\pi m^* kT)^{\frac{3}{2}}}. \tag{33}$$

Substituting $\exp\lambda$ in the expression for dn, neglecting unity in comparison with a much larger quantity, and replacing E by $m^*v^2/2$, we obtain

$$dn = 4\pi n_0\left(\frac{m^*}{2\pi kT}\right)^{\frac{3}{2}} v^2\exp\left(-\frac{mv^2}{2kT}\right)dv. \tag{34}$$

Substituting the values $n_0 = 6\times10^{22}$, $m^* \fallingdotseq m_0 = 9\times10^{-28}$ gm, corresponding to electrons in silver, and $T = 300°$ K in expression (33), we obtain for $\exp(-\lambda)$ a quantity of the order of 5000, and for $\exp\lambda$, 2×10^{-4}.

At such low values of $\exp\lambda$, unity in the denominator in expression (31a) cannot be neglected in comparison with the exponential term. Rather, the exponential term, in this case, is small as compared with unity. Electrons in a metal are thus seen to obey different statistics, namely Fermi statistics, yielding quite different conclusions from those of the classical statistics of gas molecules.

If, instead of the mass of an electron we substitute the mass of a nitrogen molecule m_{N_2} equal to $4{\cdot}5\times10^{-22}$ gm and putting n_0 equal to 10^{20}, we obtain for $\exp(-\lambda)$ a quantity of the order of 10^{-3} and for $\exp\lambda$, 10^3.

In this case the term $\exp\lambda$ is large in comparison with unity and, as we have seen earlier, we obtain for dn the expression (34) which is valid for ordinary gases.

Thus, to solve the problem of whether a given system of particles obeys the laws of classical or quantum statistics (in the latter case the system is called degenerate), it is necessary to determine the quantity $\exp[(E-\mu)/kT]$. When $\exp[(E-\mu)/kT] \gg 1$ the system may be considered to be classical, but when $\exp[(E-\mu)/kT] \ll 1$ the system is degenerate.

The factor $\exp\lambda$ may have a large value not only as a result of a large volume of the mass m but also because of a small value of the

number of particles n_0 in a unit volume. Thus when n_0 equals $10^{12}-10^{18}$ as for electrons in a semiconductor, $\exp\lambda$ acquires values of $10-10^7$. Therefore, electrons in a semiconductor usually obey gas laws and classical statistics very closely.

At room temperature the degeneracy criterion for electrons, $\exp[(E-\mu)/kT]$ equal to unity or E equal to μ, gives n_0 equal to $2{\cdot}5\times10^{19}$ cm^{-3}. As follows from expression (33), $\exp\lambda$ increases with increase of temperature.

The following question may be asked: at what temperature T will the electron gas in a metal acquire the property of an ordinary gas? For this to happen, $\exp\lambda$ will have to be of the order of unity or greater. This temperature may be estimated by putting $\exp\lambda$ equal to unity in expression (33), although under these conditions the expression (33) itself is not accurate enough. The temperature at which degeneracy disappears in a typical metal, e.g. silver, is found to be equal to

$$T_k = \frac{n_0^{\frac{2}{3}}h^2}{2\pi mk} = 7\times10^4 \text{ deg K.}$$

It is, therefore, not surprising that the hypothesis of the classical electronic theory of Drude–Lorentz, attributing to electrons at normal temperatures an average energy $\frac{3}{2}kT$, equal to the energy of gaseous molecules, yields false results. Only above a few tens of thousands of degrees would electrons in a metal satisfy the usual gas laws, i.e. the electron gas would not be degenerate. In semiconductors, however, the density of the electron gas is so small that T_k is much lower and the electron gas, even at room temperature, is often non-degenerate.

Heat capacity of electrons

By carrying out the necessary calculations we shall find out that the application of Fermi–Dirac statistics to electrons in a metal eliminates the aforementioned contradictions between the classical theory and the experimental laws of specific heat.

In fact, the energy of electrons per unit volume, according to expression (31a), may be represented in the form

$$E = \int E dn = \frac{4\pi(2m^*)^{\frac{3}{2}}}{h^3}\int \frac{E^{\frac{1}{2}}\,dE}{\exp\left(\frac{E-\mu}{kT}\right)+1}.$$

At sufficiently high values of $E_{\max}/kT$ this expression may be written down approximately, to the accuracy of second-order terms,

$$E = \frac{3}{10}\left(\frac{3n_0}{8\pi}\right)^{\frac{2}{3}}\frac{n_0h^2}{m^*}+\frac{\pi^2}{2}\frac{n_0m^*k^2}{h^2}\left(\frac{8\pi}{3n_0}\right)^{\frac{1}{3}}T^2.$$

Therefore, the heat capacity per unit volume of electrons in a metal is equal to

$$c' = \frac{dE}{dT} = \pi^2 \frac{n_0 m^* k^2}{h^2} \left(\frac{8\pi}{3n_0}\right)^{\frac{1}{3}} T. \tag{35}$$

The heat capacity of the electrons contained in 1 cm^3 of silver, with n_0 equal to 6×10^{22} and T equal to 300° K is approximately 8×10^5 erg/deg, or 2×10^{-2} cal/deg. On the other hand the heat capacity of 6×10^{22} atoms of silver occupying a volume of 1 cm^3 and possessing three degrees of freedom per atom is equal to 0·6 cal/deg.

Thus the heat capacity contributed by the electrons constitutes only 3% of the heat capacity of the metal lattice. This explains why there is no appreciable difference between the heat capacity of a gram atom of metal and an insulator. At the same time, close to the absolute zero, when the heat capacity of a solid is hundreds of times smaller, the heat capacity of the electron gas in the metal begins to predominate.

Only a small fraction of the total number of electrons in the metal is responsible for its electrical conductivity and thermal conductivity.

We shall return later to the problem of why the thermal conductivity of a metal is determined by its electrons, and why the heat transfer by the vibrating atoms in the crystal lattice has no effect, when we discuss the mechanism of thermal conduction in semiconductors.

Quantum states of the electrons in an atom

The most direct data on the energy of electrons in atoms have been yielded by the study of the optical spectra of the atoms and gas molecules. This method is inapplicable to metals since the absorption and emission of light by metals extends continuously from the lowest frequencies to the ultra-violet region of the spectrum indicating the presence of an energy band of electrons in a metal extending over a few eV. A much clearer picture is yielded by soft X-ray spectra, which substantiate the electron energy level picture yielded by the quantum theory of metals. X-rays are produced when an electron returns to one of the internal atomic levels which have been vacated previously. Removal of an internal electron occurs on the absorption of an X-ray photon; the absorption of a photon of frequency ν imparts an energy $h\nu$ to the electron.

The state of an electron in a single atom or a metal is described by quantum numbers: in particular in an atom it is described by four quantum numbers, n, l, j and m, of which the first mainly determines the energy and the second the angular momentum.

Not all transitions of an electron from one level to another are possible. For such a transition it is first of all necessary for the final

state of the electron to be unoccupied. Certain relationships between the quantum numbers of the two states must also be satisfied. In particular the second quantum number l must change by unity.

The following electron states are related to the quantum number l:

s the state at $l = 0$,

p the state at $l = 1$,

d the state at $l = 2$,

f the state at $l = 3$.

The "selection rule" mentioned above permits only transitions from state p into state s or conversely, from state p into state d or from state d into state f, but does not permit a transition from one s state into another s state.

The value of the principal quantum number n is denoted by an integer which is then linked with the appropriate symbol, s, p, d or f; the states are classified as follows:

K-levels with $n = 1$, for example, $1s$ (electrons nearest to the atomic nucleus),

L-levels with $n = 2$, for example, $2s$, $2p$ (the second electron shell),

M-levels with $n = 3$, for example, $3s$, $3p$, $3d$ (third electron shell),

N-levels with $n = 4$, for example, $4s$, $4p$, $4d$, $4f$ (fourth electron shell).

The symbols K, L, M and N were employed to denote the different groups of X-rays which were observed at a time when the atomic theory, which later clarified the physical meaning of spectra, did not yet exist. It was found later that the group of spectrum lines called the K group, results from transitions of electrons from other states to states with the principal quantum number n equal to 1 (i.e. to the state $1s$), the L group of spectrum lines is caused by transitions to states with n equal to 2, i.e. to the states $2s$ and $2p$, etc.

To each principal quantum number there is an s state in which may be present two electrons with opposite spins; there are three different p states, each of which may be occupied by two electrons. When the energies of the electrons in two or more states are equal, these states are then called degenerate.

Quantum states of electrons in a metal

When a gas becomes solid or liquid the proximity of the atoms enhances their interaction. The electrical fields of the individual atoms overlap and this alters the quantum states.

We may regard each atom of a rarefied gas as a separate quantum system and, thus, within the confines of an atom, no two electrons with all four quantum numbers identical can exist. There is a similar set of quantum states for the electrons of each atom. However, a solid or a liquid has to be regarded as a single system in which all the electrons must be in different states. When N atoms come close together, the initially identical states for the individual electrons become separated, under the effect of the electrical interactions, into N close, but different, levels, which form an almost continuous band when N is large.

In such a broadening it is found that the bands corresponding to the different energy levels of an atom may partially overlap and electrons with equal energies may be in different states.

The proximity of the atoms has a different effect on the electrons belonging to K, L, M and N levels. For the outer valence electrons the energy of interaction with adjacent atoms is of the same order or even larger than the energy of interaction with the nucleus of their own atom and, therefore, the entire set of valence electrons forms a single system associated with the whole body, there being no bond between such an electron and a specific nucleus.

The position becomes different for the case of the inner electrons; here the energy of binding with their own nucleus, measured in hundreds and thousands of electron volts is much higher than the energy of interaction with a neighbouring nucleus which does not exceed a few electron volts. The wave functions of the inner electrons hardly interact and, therefore, the states of each electron remain virtually unchanged. Although these electrons may be considered as lying in single systems they may also be regarded as forming a part of a single atom without much error.

X-ray spectra of metals

The very small magnitude of the distortion of the wave functions of the internal electrons facilitates the interpretation of the X-ray spectra of metals. The internal electron levels (the K-level and L-level) are almost the same in a metal as in a free atom and the energy is almost as sharply defined. The ranges of energy change for electron transitions, therefore, indicate the widths of the energy bands from which the electrons pass to the given K- or L-level.

We shall now find out what is to be expected on the basis of quantum statistics. Any one of the valence p-electrons may pass to an unoccupied K-level; the probability of transition increases with frequency ν being proportional to ν^3. Since, even at the absolute zero, the electrons in a metal occupy states with energies from zero to E_{max} of about 5–10 eV, it may be expected that the energies liberated by valence electrons on

transition to the K- or the L-level and, consequently, also the energies of the X-ray photons, will range from a certain E_0 to (E_0+E_{max}), whilst their frequencies will occupy a certain continuous spectral band with a width E_{max}/h. The distribution of n electrons within the energy range 0 to E_{max} should correspond to the energy distribution $J(E)$ within the spectral band, i.e.

$$n(E) \propto \frac{J(E)}{\nu^3}.$$

The results of the experimental study of X-ray spectra emitted by various metals have on the whole confirmed these predictions. The spectra contain bands, the width of which $(\nu_m - \nu_0)$ corresponds to the band of energy levels of electrons $(\nu_m - \nu_0) = E_{max}/h$. It would be expected that the dependence of n on E, at the absolute zero, should be expressed by the parabola $n \propto E^{\frac{1}{2}}$ in accordance with expression (23). From the point of view of the classical theory of metals which assumes that the energies of the electrons, like the energies of gas molecules, are equal to about $3kT/2$, the X-ray spectra would have been expected to yield relatively narrow lines with a width of approximately 0·03 eV.

The dependence of the number of the electrons on their energy

$$n \propto \frac{J(E)}{\nu^3} \propto E^{\frac{1}{2}}$$

which follows from expression (23) is only roughly confirmed by experiments for the electrons which are initially in an s state and which pass into a p state, corresponding to L_3 levels (Figs. 16 and 17).

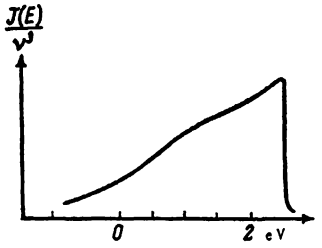


Fig. 16. The shape of the emission spectrum L_3Na.

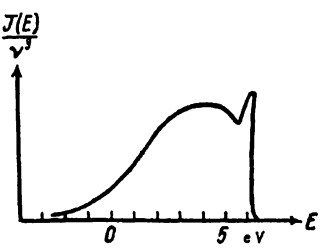


Fig. 17. The shape of the emission spectrum L_3Mg.

In Fig. 16 the line L_3Na corresponds to the transition from the $3s$ band and into the $2p$ state. This curve gives the energy distribution of electrons in the band of $3s$ states of solid Na. The ratio $J(E)/\nu^3$ is a measure of $n(E)$.

On the other hand, spectra corresponding to the transition of electrons from a p band to the s state at the K-level exhibit a steeper rise of n with increase of E, namely, according to the law $n \propto E^{\frac{3}{2}}$.

The difference between the properties of the s and p electrons has not been taken into account in the simplified theory of metals outlined above. The physical meaning of this difference will be clarified in the more complete band theory of solids. The experimental values of the Fermi level energy are, however, in good agreement with the values calculated from expression (25).

TABLE 4

Metal	No. of valence electrons per atom	Experimental E_{max} (eV)	Calculated E_{max} (eV)
Li	1	$4{\cdot}1 \pm 0{\cdot}3$	4·8
Be	2	$14{\cdot}8 \pm 0{\cdot}5$	14·6
B	3	$29{\cdot}5 \pm 2$	25·2
Na	1	$3{\cdot}0 \pm 0{\cdot}2$	3·2
Mg	2	$7{\cdot}3 \pm 1$	7·3
Al	3	$12{\cdot}7 \pm 1$	11·9

The values of E_{max} obtained from X-ray spectra are compared in Table 4 with the calculated values on the assumption that all the valence electrons occupy the lowest quantum energy levels.

The data in Table 4 leave no doubt that the energies of electrons in a metal are determined not by a statistical spread about an average value kT equal to 0·03 eV, but by a quantum statistical law, the valence electrons in a metal having a wide energy distribution.

The quantum theory shows how, when going over from the absolute zero to increasingly higher temperatures, the distribution of electrons among energy levels is changed. Only changes within about kT of the upper energy limit (the Fermi level) are possible. The steep decline of the curve $n(E)$ with the rise of T in the vicinity of the absolute zero becomes more gentle; at the same time the number of electrons with energies below the Fermi level decreases and an increasing number of the electrons pass into states with energies exceeding E_{max}. This conclusion is fully substantiated by X-ray spectra. Fig. 18 shows the intensity distribution $J(E)$ of the X-ray spectrum band of magnesium close to its upper edge, corresponding to the Fermi level, at temperatures of 110, 300 and 670° K. Division of $J(E)$ by ν^3 would yield the energy distribution of the electrons. The position of the Fermi level here corresponds to the centre of each of the three curves.

The aforementioned deviation of the theoretical shape of the curve $J(E)/\nu^3$ equal to $f(E)$ from the experimental shape shows that in spite

of the general confirmation of the theory, its underlying hypothesis of free electrons is inadequate; it is certainly inadequate when dealing with the band of levels in which the various quantum states overlap as in the case of magnesium.

One could next ask whether, in accordance with our hypothesis, the electrons in a metal are in fact characterized only by their energies or if they retain the properties of the quantum states of the atom which they occupied before condensation of the vapour to the solid metal. Investigation of the X-ray spectra supports the latter hypothesis. It is found that the same selection rules are also valid for the electrons in the metal.

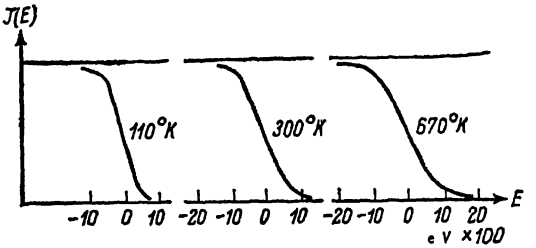


Fig. 18. Broadening of the edge of the emission spectrum L_3Mg with rise of temperature.

As in optical transitions in an atom, for example to the $1s$ levels (which correspond to the K-line of the spectrum), there pass only electrons which are in the p states, and on to the L_3 band, corresponding to the $2p$ level, there pass only electrons from the s and d states. Therefore the K and L spectra of one and the same metal differ from each other and portray the energy distributions of different electrons.

The assumption that the energy distribution of electrons in a metal is independent of the crystallographic orientation may hold for a perfectly isotropic solid and is also applicable to a certain degree of liquid metals. Solid metals, however, possess a crystalline lattice the properties of which are different in different crystallographic orientations. Some metals (Bi, Zn, etc.) are so anisotropic that the energy level distribution in them is far from spherical and differs very greatly for different orientations.

Finally, it should be remembered that when the atoms come closer together the distribution of quantum states of atomic electrons may change. Thus, for example, the valence electrons in the divalent atoms of Be, Mg, etc., which in individual atoms possess opposite spins, pass into states with spins directed the same way during the formation of molecules.

Therefore, the quantum theory of metals outlined above should be regarded as a rough approximation of a much more complex theory of real metals. The development of this theory will be considered in conjunction with the theory of electrons in semiconductors, the properties of which depend substantially just on those factors which have not been taken into account in the approximate theory of free electrons.

Distribution of electron velocities in an electrical field

An electrical current arising in a metal under the effect of potential difference finds its expression in the predominance of velocities in the direction of the field over velocities in the opposite direction.

We have seen, however, that at the absolute zero all possible states with energies below the Fermi level are occupied by the electrons, and at higher temperatures only a small fraction of the states are vacated at a distance of the order of kT below this level. All the electrons occupying low-energy states more distant from the Fermi level cannot change either the magnitude or the direction of their velocity, since all the states with close energy values are occupied by other electrons. Only the small fraction of electrons in a metal with energies close to the Fermi level can change their states under the action of a field; the velocity component of these electrons directed along the field increases, and the component in the opposite direction decreases.

An electrical field $\mathscr{E}$ accelerating these electrons between collisions imparts to them an additional velocity Δv

$$\Delta v = \frac{1}{2} \frac{\mathscr{E} e}{m^*} \frac{1}{v}$$

where v is the velocity of electrons with energies close to the Fermi level.

Electrons moving along the field increase their velocity by this amount whilst electrons moving against the field are slowed down by the same amount. Therefore the energy distribution of electrons close to the Fermi level, which did not depend on the velocity direction before the application of the field, changes when the field is applied: the energy of the fastest electrons increases by an amount corresponding to their acceleration by the field. Electrons with lower energy will now be able to occupy the sites vacated by the fast electrons. As a result of this, the entire velocity distribution pattern will shift in the direction of the field by an amount equal to the velocity increment of the fastest electrons.

Fig. 19 shows the distribution of velocity components of the electrons (solid line): on the positive side along the direction of the field, and on the negative side in the opposite direction. In the absence of an electrical

field, the velocity distribution is symmetrical and the solid line represents the velocity projections in an arbitrary direction. An electrical field may accelerate or slow down the motion of only those electrons which are at the boundary with the free states. These conditions exist among the electrons which are close to the Fermi level. Under the effect of the field the velocities of these electrons change. The broken line shows the new distribution under the effect of the electric field.

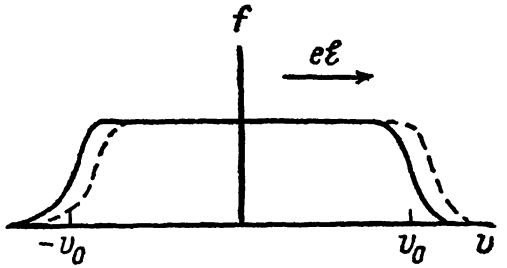


Fig. 19. Velocity distribution of electrons in a metal.

The new distribution differs from the earlier one only by the addition of a certain number of electrons with velocities exceeding the Fermi velocities and the disappearance of a small number of electrons with maximum velocity with a direction opposing the field; the remainder of the graph remains unchanged. All the same, the result of action of the field is exactly the same as that which would be produced by shifting the velocity distribution of all the electrons to the right by a certain amount which is proportional to the field strength.

Individual electrons are indistinguishable. We shall therefore consider, not the shift of each individual electron, but the transition of all the electrons from a certain steady state in the absence of the field to the new steady state in the electrical field. In doing this we refrain from describing the kinetics of the process which sets up the new equilibrium and confine ourselves to the consideration of its results. The result is such that the n_0 electrons in the metal now have a velocity component along the direction of action of the field larger by an amount

$$\Delta v = u\mathscr{E}$$

and the current density j produced by the field is equal to

$$j = n_0 eu\mathscr{E}.$$

The factor u may be regarded as an average mobility of electrons in a metal, i.e. an average additional velocity acquired by all the free

electrons n_0 in a field of unit strength, although in reality only electrons close to the Fermi level participate in the current and in the heat transfer.

Thermionic emission

In a metal at 0° K the electrons possess kinetic energies ranging from zero to E_{max}. As the electrons do not leave the metal, there must obviously exist electrical fields close to the surface which return those electrons with velocities directed outwards. Since at temperatures higher than the absolute zero there are always some electrons with energies exceeding E_{max}, it is necessary to assume that the energy required by electrons to leave the metal is much greater than E_{max}. The excess energy ϕ above E_{max} which must be imparted to an electron to allow it to pass from the metal into the surrounding space (assuming that the metal is *in vacuo*) is called the work function. Its value is a few electron volts and may be expressed as $\phi = eV_c$ where V_c is the contact potential of the given metal and e the electronic charge.

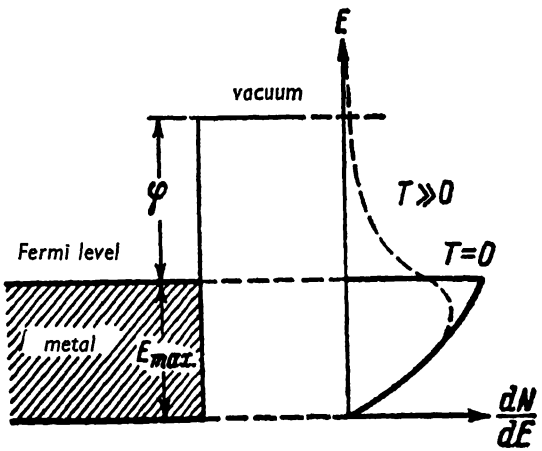


Fig. 20. Energy diagram of electrons in a metal. On the right-hand side is shown the energy distribution of electrons at $T = 0°$ K and $T \gg 0°$ K.

Fig. 20 shows an energy diagram for electrons in a metal. Taking the average potential energy of an electron inside a metal to be equal to zero, we can say that the energy of an electron outside a metal is equal to $E_{max} + \phi$.

Strictly speaking, at all temperatures above the absolute zero there may exist a certain proportion of electrons with energies exceeding E_{max} by ϕ and more. The number n of such electrons is usually negligible.

According to equation (31) it is given by the expression

$$n = \frac{4\pi(2m^*)^{\frac{3}{2}}}{h^3} \int_{E_{\max}+\phi}^{\infty} \frac{E^{\frac{1}{2}}\,dE}{\exp\dfrac{E-\mu}{kT}+1},$$

or to a sufficiently good approximation

$$n = \frac{4\pi}{h^3}(2m^*\,kT)^{\frac{3}{2}}\left(\frac{E_{\max}+\phi}{kT}\right)^{\frac{1}{2}} \exp(-\phi/kT). \tag{36}$$

The thermionic current from unit surface area of the metal is

$$j = \frac{4\pi em^*\,(kT)^2}{h^3} \exp(-\phi/kT).$$

Since ϕ is equal to a few (3–5) electron volts and kT at room temperature amounts to only 0·03 eV, the factor $\exp(-\phi/kT)$ is negligibly small (see Table 1). For ϕ equal to 2 eV, $\exp(-\phi/kT)$ is equal to only 10^{-30}, and therefore the probability that 1 cm^3 of metal will contain even one electron capable of leaving it for one second is negligibly small. However, the number of such electrons grows steeply with rise of temperature and, for example, at a temperature of 1200° C or around 1500° K the value of kT becomes equal to 0·15 eV, and $\exp(2/0{\cdot}15)$ is equal to 10^{-6}. At this temperature with ϕ equal to 2 eV a substantial proportion of the electrons (there are approximately 10^{15} electrons in 1 cm^2 of the surface layer) will be capable of leaving the confines of the metal.

This phenomenon is in fact observed at elevated temperatures. It is named thermionic emission. The velocity distribution of electrons leaving the metal is in good agreement with the quantum theory but at the same time it does not differ from that predicted from the classical theory.

When two different metals are brought into contact, the electrons pass from the metal where their energies are higher into the metal where the energies are lower, imparting to the latter a negative charge; the electrostatic energy of the electrons which have undergone such a transition increases. The transfer of the electrons continues until the total energy, i.e. the sum of the potential and kinetic energies, becomes equal in both metals. The condition of thermal equilibrium in solids consists of the equalizing of the free energy. As we shall see later, in the case of metals, this signifies equal Fermi levels (Fig. 21b).

The potential difference V_c between the metals which is set up on attaining the equilibrium is called the contact potential difference. It sets up an electrical field $\mathscr{E}_c$ in the space between the metals (Fig. 21a).

The contact potential difference V_c is determined as the difference between the energies ϕ_1 and ϕ_2 which must be imparted to electrons at the Fermi level to allow them to pass through the surface of the metals into the surrounding medium, e.g. the vacuum. The energy change cannot depend on the path along which the energies are transferred and $(\phi_1-\phi_2)$ is the same in the case of the electrons passing through the surrounding medium as in the case of the direct transition of electrons from one metal into another through the interface

$$\phi_1-\phi_2 = V_c e.$$

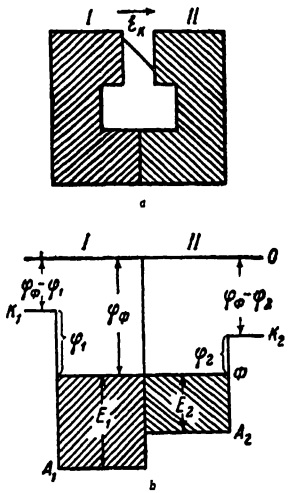


Fig. 21. Contact between two metals (I, II): (a) field in the gap between the metals; (b) energy diagram illustrating the formation of the contact potential difference.

The line marked O in Fig. 21b denotes the energy level of the surroundings assumed to be equal to zero. When measured from this reference line the energy of electrons at the Fermi level ϕ_Φ is equal for both metals (I and II). Inside the metal the energy of the electrons is negative with respect to the surrounding medium and therefore the electrons are retained in the metal. The levels K_1 and K_2 denote electron energies sufficient for the electrons to leave the metal. Their energies with respect to the Fermi level are equal to ϕ_1 and ϕ_2 and, with respect to the zero level, $(\phi_\Phi-\phi_1)$ and $(\phi_\Phi-\phi_2)$ respectively. Since the electronic charge e is negative, the contact potential difference between the metals which have attained equilibrium will be equal to

$$V_c = \frac{(\phi_\Phi-\phi_1)-(\phi_\Phi-\phi_2)}{-e} = \frac{\phi_1-\phi_2}{e}.$$

The lower straight lines A_1 and A_2 denote the energies of electrons with the minimum (zero) kinetic energies. Their distances E_1 and E_2 from the Fermi level are equal to the kinetic energies of the electrons at the Fermi level. The sums $(\phi_\Phi + E_1)$ and $(\phi_\Phi + E_2)$ give the potential energy of the electrons in the two metals acting to prevent the electrons from leaving the metals. The difference between the potential energies $(E_2 - E_1)$ is equal and opposite in sign in the difference of kinetic energies at the Fermi level, since the sum of potential and kinetic energy of the electrons at this level should be equal in both metals after the establishment of equilibrium.

Electrons corresponding to energy levels K_1 and K_2, on leaving the metal perform work against the forces of attraction towards the metal equal to $(\phi_1 + E_1)$ or $(\phi_2 + E_2)$, i.e. equal to the sum of the kinetic energy E at the Fermi level and the extra kinetic energy ϕ. The kinetic energy of such electrons which have left the metal is found to have gone entirely into detaching the electrons from the metal and thus drops to zero. On the other hand, the electrostatic energy which sets up a contact potential difference V_c between different metals is determined by the difference between the distances of K_1 and K_2 from the zero energy level.

Electrons within the metal which possess kinetic energies exceeding $(E_1 + \phi_1)$ or $(E_2 + \phi_2)$ may, on leaving the metal, retain their excess energy when they have passed through the metal boundary in a direction normal to the surface. When the electrons move at an angle to the surface only a portion of the kinetic energy can be used to overcome the work function; the velocity component parallel to the surface of the metal remains unchanged since it is perpendicular to the force acting on the electron. The quantum theory shows that not all electrons possessing the necessary energy leave the metal. Some are reflected from the boundary and return into the solid after changing the direction of their velocity.

The excess energy which the electrons require for passing through the surface of the metal may be imparted to the electrons not only by thermal agitation but also by other sources of energy. When the source consists of photons we observe photoelectric emission. To enable an electron which has absorbed the energy of a photon to leave the metal, the sum of the energy $h\nu$ imparted by the photon and the kinetic energy of the electron E must exceed the value $(E_{max} + \phi)$

$$(E + h\nu) \geqslant (E_{max} + \phi). \tag{37}$$

The bulk of the electrons detached from the metal by light possess energies E not exceeding E_{max}. Therefore condition (37) coincides with

the requirement

$$h\nu > \phi \tag{38}$$

which determines the minimum frequency of the light ν_0 which can produce an external photo effect (the red edge of the photo effect). Fig. 22 illustrates the dependence of the photoelectric current on light frequency.

When the excess energy is imparted by fast electrons penetrating the metal from the outside and giving up part of their energy to electrons in the metal, the phenomenon of electron emission is called secondary electron emission.

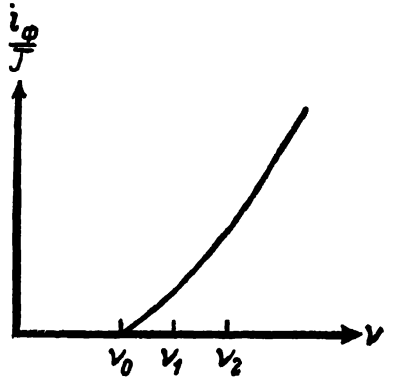


Fig. 22. Spectral distribution of the normal external photoeffect: i_Φ, photoelectric current; J, radiant energy flux; ν_0, limiting frequency for the photoeffect.

Magnetic properties of electrons

The magnetic susceptibility of metals is composed of three parts: (1) the magnetic susceptibility of the metallic atoms (paramagnetic or diamagnetic depending on their chemical nature); (2) the paramagnetism of conduction electrons, caused by the predominance of electrons with a spin directed along the field over electrons with an opposite direction of the spin; (3) the diamagnetism of electrons moving in the metal in random directions.

The probability that the spin of an electron in a given quantum state will be directed along the magnetic field H is equal, according to equation (30a), to

$$p_+ = \frac{1}{1+\exp\left(\frac{E_0-\mu}{kT}-\frac{ehH}{2m^* ckT}\right)} \tag{39}$$

where $eh/2m^* c$ is the magnetic moment of an electron due to its spin and E_0 the energy of an electron with no field.

For the probability of the opposite direction of the spin we obtain

$$p_- = \frac{1}{1+\exp\left(\frac{E_0-\mu}{kT}+\frac{ehH}{2m^* ckT}\right)}. \tag{39a}$$

Adding up the values $(p_+ - p_-)$ over all electron states, we find the excess number of electrons with a spin directed along the field over the number of electrons with an opposite spin. Attributing to each electron the magnetic moment $\pm eh/2m^* c$, we obtain the magnetic moment M of all electrons, and dividing M by H the paramagnetic susceptibility χ_+.

We find approximately

$$\chi_+ = \frac{e^2}{4\pi m^* c^2}\left(\frac{3}{\pi}n_0\right)^{\frac{1}{3}}. \tag{40}$$

The diamagnetic susceptibility χ_- of free electrons given by the classical theory of metals is equal to zero. Landau has shown, however, that the quantum theory leads to the value

$$\chi_- = \tfrac{1}{3}\chi_+.$$

The above calculations are based on the assumption that individual electrons are independent of each other. Interaction between the electrons under certain conditions leads to the formation of domains with all spins in the same direction (ferromagnetism) or with pairs of spins in opposite directions (anti-ferromagnetism).

Thermoelectric phenomena

In 1821 Seebeck discovered the phenomenon which was found later to consist of the appearance of an e.m.f. dV between two conductors with junction temperatures differing by dT

$$dV = \alpha dT$$

where α is thermal e.m.f. coefficient.

Thirteen years later Peltier discovered another phenomenon closely related to the first one. A current J passing through the junction between two different conductors causes a certain amount of heat energy Q to be liberated or absorbed at the junction depending on the direction of the current; the power q equal to Q/t is proportional to the current

$$q = Pj.$$

The Peltier coefficient P is related to α by the following relationship

$$P = \alpha T$$

where T is the absolute temperature of the junction between the conductors.

Both phenomena are explained by the dependence of the state of the electrons in a conductor on temperature causing the diffusion of electrons from the hot end to the cold end and from one conductor to another conductor in contact with it.

These phenomena also occur in ionic conductors.

As opposed to the case of semiconductors, the Fermi level in a metal is virtually independent of temperature: a change of temperature leads to a change of the state of a small number of electrons the energy of which is close to the Fermi level. Consequently, thermoelectric phenomena in metals are much less pronounced. When only linear terms in the expansion of the quantity α are taken into account, this quantity is found to be equal to zero. Higher order terms give a value for the thermal e.m.f. α of simple metals of

$$\alpha \simeq \pi^2 \frac{k}{e} \frac{kT}{E_\Phi}$$

where E_Φ ($= E_{max}$) is the energy of electrons at the Fermi level. At room temperature kT/E_Φ has a value of the order of 0·002; $\pi^2 k/e$ is about 9×10^{-4} V. Therefore in metals the value of α is of the order 2×10^{-6} V.

In fact, the thermal e.m.f. of the majority of pure metals does not exceed a few μV per degree. In accordance with the expression for α it rises proportionally to the absolute temperature.

However, for lithium α is expressed in tens of μV/deg. It has been shown by Samoilovich and Serova[1] that this exception is due to the fact that in the case of lithium there exist conditions under which the electrons are dragged by the heat flux from the cold end of the conductor to the hot end, the phenomenon predicted by Gurevich.[2,3]

In some alloys α reaches values of up to 50 μV/deg. Such alloys (constantan, chromel, alumel) are used for the measurement of temperature.

Let us note, without yet giving any explanation, that among the metals there are some in which the potential at the hot end is higher than at the cold end (Na, K) and some for which α has the opposite sign (W, Mo, Zn), and finally that there are metals (Sn, Pb) for which α is close to zero. In all these cases the sign of α corresponds to the sign of the Hall effect.

The physical significance of these facts will be considered in conjunction with the theory of semiconductors for which the thermal e.m.f.s have much higher values.

References

GENERAL

Ya. I. Frenkel', *Vvedenie v teoriyu metallov* (*Introduction to the Theory of Metals*), 2nd ed. Moscow-Leningrad, Gos. Izd. Tekh.-Teor. Lit., 1950.

F. Seitz, *The Modern Theory of Solids*. New York-London, 1940.

H. Fröhlich, *Elektronentheorie der Metalle* (*Electron Theory of Metals*). Berlin, J. Springer, 1936.

A. Sommerfeld and H. Bethe, "Elektronentheorie der Metalle" (Electron theory of metals), *Handbuch der Physik*, 2nd ed., Vol. XXIV/2. Berlin, J. Springer, 1933.

SPECIFIC PROBLEMS

1. F. G. Serova, "Thermoelectric properties of monovalent metals", *Zhur. Eksper. Teor. Fiz.*, **19**, 460 (1949).
2. L. E. Gurevich, "Thermoelectric properties of semiconductors—I", *Zhur. Eksper. Teor. Fiz.*, **16**, 193 (1946).
3. L. E. Gurevich, "Thermomagnetic and galvanomagnetic properties of semiconductors", *Zhur. Eksper. Teor. Fiz.*, **16**, 416 (1946).

CHAPTER

3

Electronic Semiconductors—A General Survey

Types of semiconductor

Substances can be divided into types according to the nature of the forces that bind the atoms together, namely:

1. Those in which the atoms are bound by valence bonds, each formed in a definite crystallographic direction by a pair of electrons. This type of bond causes the crystal to be hard and strong; diamond, silicon, germanium, grey tin and silicon carbide have bonds of this type. Each atom has a number of nearest neighbours (the coordination number) that equals its valency, and in the above cases the number is four.

2. Ionic crystals, typified by rock salt (sodium chloride), which consist of positive and negative ions bound by Coulomb forces between the charges.

3. Molecular crystals (usually of organic compounds); the molecules are bound together by rather weak Van der Waals forces. Such crystals are usually soft and have low melting points.

4. A metal has a special type of bond, namely that between its assemblies of free valence electrons and positive ions. The strength and plasticity of metals result from this bond.

These four types are seldom found in pure form, but usually occur in combinations. Thus, if we first suppose sodium chloride to have purely ionic bonds, we must assume that the eleventh electron from a sodium atom is transferred entirely to a chlorine atom. However, we cannot even then take each ion as being strictly independent. The Coulomb forces between the charges polarize the chlorine ion and draw its electrons over towards the sodium ion, which effect must be reflected in the bond strength. The bond becomes partly of the valence type. Neither can we neglect the interactions between the ions caused by the periodic motion of the electrons; forces of the Van der Waals type result.

To assume a pure valence bond would imply that two electrons are equally shared between the atoms; this cannot occur precisely if the

atoms are different so that the bond becomes partly ionic. In crystalline selenium and tellurium there are valence bonds between the atoms in any one chain, and Van der Waals forces between chains.

It is also possible to quote instances where ionic or valence bonds between separate atoms are combined with metal-type bonds between the ions and the electrons.

Semiconductors can be either crystalline or amorphous; instances of the latter type are liquid semiconductors and supercooled liquids in the vitreous state, e.g. solid amorphous selenium. Crystalline materials are often met in the form of polycrystalline aggregates, pressed or sintered granular masses, or thin layers deposited on sub-strata. The properties of any such material depend on its structure.

By semiconductors we usually mean non-metallic conductors with an electronic conduction mechanism. We have pointed out earlier that it is often not easy to find out the precise conduction mechanism. It will be simpler to see how to solve this problem when we have dealt with the main similarities and differences between semiconductors, electrolytes and metals.

Comparison with electrolytes

Electronic semiconductors are like solid electrolytes in that in both cases the carriers (electrons in the first, ions in the second) are freed by thermal agitation. The free-charge concentrations, and the conductivities, thus tend to zero as the absolute zero of temperature is approached. Free electrons diffuse through a body for a certain time like ions before they are trapped at lattice defects, where they may recombine. Here we find a difference, for, whereas ions usually recombine by falling into vacant lattice sites, electrons usually recombine at impurities or at lattice defects. An accumulation of ions of the same sign, or of electrons, produces a space charge. The two types of conductor differ entirely in the mechanism which determines the carrier mobility, however. Ions have to surmount potential barriers in order to pass from one unit cell to another, whereas electrons pass freely through the potential barriers between atoms.

The relation between the size of the ion and the packing density of the medium in which the ion must move is decisive for ionic motion. There are thus no great differences between the mobilities of ions in a highly perfect single crystal and in an amorphous solid; the ions pass more freely if the lattice becomes more loosely packed. Ionic mobilities increase rapidly with temperature, and are higher in melts than they are in solids.

Electrons would move without any resistance in the periodic field of an ideally perfect crystal; deviations from strict periodicity alone

can cause scattering. Hence electron mobilities usually fall as the temperature rises, when the lattice distortions increase, or as impurities are added; they also become much smaller if the substance is in the liquid or amorphous states.

The ions in electrolytes all have the same mean kinetic and potential energies, whereas the electrons in a semiconductor are spread over energy levels that have a range of about 1 eV.

Comparison with metals

A semiconductor has a small value of σ not because u, the mobility, is lower than in a metal, but because only a few of the electrons carry the current. For example, a semiconductor may have σ equal to 10^{-3} ohm^{-1} cm^{-1}, u equal to 100 cm^2/V sec and a free electron concentration n equal to 10^{14} cm^{-3}, whereas the total valence concentration is about 5×10^{22} cm^{-3}. The mobility increases as the semiconductor is cooled, although the conductivity falls. Thus the fall in σ must be caused by a very rapid fall in n. Since σ and n both tend to zero as $T \rightarrow 0$, we consider that thermal agitation produces the conduction electrons. Here we have the main great difference between semiconductors and metals, though there are some materials in which it is difficult to distinguish between metallic and semiconducting behaviour.

The electron waves are scattered in semiconductors for the same reasons as in metals; in metals the wavelengths are of the order of the atomic spacings, and so the waves are scattered mainly by the fastest lattice vibrations, whereas in semiconductors the wavelengths are some 20 times greater and the waves are scattered primarily by slower lattice vibrations which are much fewer in number. Moreover, all the ions in a metal have positive charges, whereas many semiconductors have either ions of both signs (ionic lattices) or neutral molecules (atomic and molecular lattices). It is clear, therefore, that the scattering, and hence the mobility u (and the way u varies with temperature), may differ for metals and semiconductors.

The following types of thermal wave motion occur in an ionic crystal:

1. Longitudinal and transverse waves, in which adjacent positive and negative ions move in the same sense. These acoustic waves cause no alternating electric fields, and scatter electron waves relatively weakly.

2. Transverse waves in which adjacent positive and negative ions move in opposite senses. These are the waves that are responsible for the frequency of the infra-red residual rays. They cause electric fields, but they do not produce groups of positive and negative charges of sizes which are comparable with the electron wavelengths (10^{-6} cm). They also scatter only weakly.

3. Longitudinal waves that make the charges move in opposite senses, with the result that in some places the charges are farther apart and in others closer together, cause the greatest scattering. These longitudinally polarized waves give the largest local field distortions.

Thermal fluctuations in lattices consisting of neutral atoms distort the periodic field, it is true, but these distortions scatter the electron waves much less than do the polarization waves in ionic lattices. The acoustical waves differ in frequency from the polarization waves, and thus the conditions under which the waves exchange energy with the lattice also differ. In all cases so far considered the mobilities rise with temperature, although different laws are followed.

Impurity ions also scatter the electron waves (and so affect the mobility), because their charges produce electric fields, although these fields are weakened by a high dielectric constant. The impurity ions scatter slow electrons more readily than fast ones; the mean kinetic energy of an electron in a semiconductor is proportional to the absolute temperature, and therefore the ions scatter less strongly as the temperature rises. This allows the mobility to rise with temperature.

The conditions in a semiconductor differ from those in a metal in yet another way. In a metal the free electrons are equal in number to the atoms, and interact with one another very strongly. The assembly must be considered as a single system that obeys Fermi statistics, and thus most of the current-carrying electrons in a metal have energies in the 5–10 eV range. In a semiconductor the number of atoms exceeds the number of free electrons by many orders of magnitude; adjacent free electrons are tens or hundreds of atomic diameters apart. The electrons interact mainly with the lattice; electron–electron interactions can often be neglected. Fermi statistics still apply to the electrons bound within the atoms. The few free electrons are like a gas, and have a mean thermal energy equal to that of the lattice; this is not so for electrons in a metal. The degeneracy factor $\exp[(E-\mu)/kT]$, applied to a semiconductor, has a value much greater than one; hence for the electrons we can use classical statistics, which also apply for the molecules of a gas.

The electrons in a semiconductor have mean thermal energies of $3kT/2$, as do gas molecules, but the electrons in metals do not. The value of this quantity at room temperature is about 0·027 eV, which is some 300 times smaller than the limiting energy in a metal. The electron wavelength is, accordingly, about 10^{-6} cm (20 times larger than for a metal). This is the reason why electrons in a semiconductor are scattered by disturbances 10^{-7} to 10^{-6} cm in size, and not by imperfections of atomic dimensions.

Sign of the current carriers

A mobility can be measured in terms of the conductivity σ and the Hall constant R. It is found that R varies in both sign and magnitude in metals and in semiconductors. The variations in magnitude of R indicate (see equation (13)) that n can vary widely; the variations in sign show that the carriers can bear either positive or negative charges. We know only of negative electrons; cosmic rays and nuclear radiations produce positrons, but energies of at least 10^6 eV are needed to produce them. It is quite impossible for thermal vibrations of energy 0·03 eV to produce positrons.

We will consider the mechanism of the Hall effect in more detail, because the fact that negative and positive charges both carry the current is very important.

The basic laws of electromagnetism deal with the force on a current in a magnetic field. The left-hand rule indicates that a conductor carrying a current from left to right, in a magnetic field directed away from us, is deflected upwards, no matter what the mechanism whereby the current is carried. The effect is not caused by the field acting on the conductor directly; the field affects the motion of the charges, and causes them to move at right angles to the field and to the direction of the conductor. It is of no importance whether positive charges are moving from left to right, or negative charges from right to left. The charges collide with impurities in the lattice, or are scattered by thermal fluctuations, and so transfer part of their momentum to the conductor, which transfer causes the conductor to move upwards.

The charges are deflected by the magnetic field, so an electric field is set up between the upper and lower parts of the wire; this field is just sufficient to balance the effect of the magnetic field. The field can be calculated if we assume that the charges move as in a vacuum, and that they all have the same speed v. We shall see later whether these assumptions are justified.

A field H acting on a charge e moving along a wire at a speed v gives a force F_1 equal to evH/c; the electric field $\mathscr{E}_H$ produces an equal and opposite force $F_2 = e\mathscr{E}_H$. The equilibrium condition is thus

$$e\mathscr{E}_H = \frac{evH}{c}.$$

We multiply both sides by n (the number of charges per cm^3) and get

$$ne\mathscr{E}_H = \frac{nevH}{c}.$$

But nev is the density j of the current J

$$j = \frac{J}{S}$$

where S equal to bd is the cross-section area of the wire, taken as rectangular (of width b and thickness d). Then $\mathscr{E}_H$ is given by

$$\mathscr{E}_H = \frac{V_H}{d}$$

where V_H is the Hall potential difference. Hence

$$V_H = \frac{JH}{necb}.\dagger$$

We again get equation (12), in which b was taken as unity, while $1/nec$ was, from equation (13), denoted by R. We see that this explanation of the Hall effect shows that the carriers charge the conductor in those regions into which the magnetic field deflects them. The sign of the resulting charge is the sign of the carriers themselves.

The thermal e.m.f. gives us another way of finding the sign of the carriers. The concentration and thermal energies of the free carriers must be larger at the hot junction than they are at the cold. The carriers diffuse to the cold junction, and of course take their charge with them. The cold junction becomes negatively charged if the carriers are electrons, or positively charged if the carriers are some kind of free positive charge.

It is found that some semiconductors have positive carriers, others negative ones, and some both types at once. The Hall effect and the thermal e.m.f. usually give the same result, which shows that carriers of both signs can exist. Metals may also have positive and negative carriers. We now examine how it happens that we can have carriers of both negative and positive sign.

If the charges in a solid behaved as do charges *in vacuo*, we should be forced to conclude that positive electrons exist as well as negative ones, but this conclusion would be wrong. The limits, to which the similar behaviour of electrons *in vacuo* and in a solid extend, and the differences, are defined by the quantum theory of solids, which we shall consider in the next chapter. In particular, we shall see that electrons cannot move with certain speeds in certain crystallographic directions, and that some energies are forbidden altogether; in the latter fact we find a likeness to the electrons in an isolated atom. The

† If we allow for the fact that v shows a statistical distribution, we find that a further factor appears on the right; this factor is $3\pi/8 = 1{\cdot}17$ for lattices in which the mean free path is independent of v.

conduction mechanism that gives a positive Hall effect can be understood correctly only from quantum mechanics.

So far we have not gone into details. A real semiconductor often has impurities; if these are of such a type, and are so placed in the lattice, that they may easily lose electrons that can then move freely, we have a conductor with free negative electrons. There are also impurities that take up electrons from the lattice, which then becomes deficient in electrons; the impurities form negative ions. The vacancies, or holes, as they are called, which are left behind are not bound to specific atoms, and diffuse through the crystal as do free electrons.

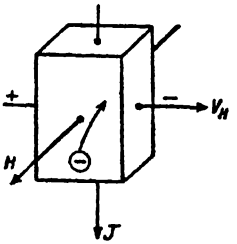


Fig. 23. The Hall effect in an *n*-type (electron) semiconductor.

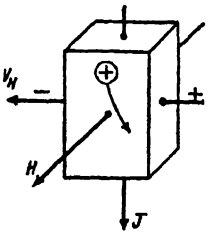


Fig. 24. The Hall effect in a *p*-type (hole) semiconductor.

A hole diffuses because a nearby electron falls into it and so leaves its own place empty. The hole migrates in a sense opposite to that for electron motion. An electric field that moves the electrons from left to right will move the holes from right to left, i.e. will move the holes as though they were positive charges. It is simplest to study the motions of the electrons that fall into the holes in terms of the motion of the holes themselves. A hole moves as though it is an electron with a positive charge.

A semiconductor with a hole conduction mechanism is said to be a *p*-type semiconductor, whereas one with an electron mechanism is said to be an *n*-type semiconductor.

If the lattice itself provides the electrons, every electron so freed leaves behind it a hole; hence hole and electron conduction mechanisms occur together.

The holes drift on an average along the direction of the electric field; they thus behave as though they are electrically positive. We cannot say *a priori* whether the holes have the same mobility as do free electrons. However, we can establish the sign and magnitude of the mobility from the Hall effect. Figs. 23 and 24 show how the paths of electrons and of holes are deflected, and the sign of the resulting e.m.f.s.

It can happen that holes and electrons both carry the current; the Hall effect then depends on the difference between the two currents, whereas the conductivity depends on the sum. Equations (13) and (14) have then to be replaced by others. The Hall effect can be zero, even though the electrons and holes both have large mobilities.

Many technical semiconductors (e.g. cuprous oxide, selenium) have pure hole conduction, and are p-type; other semiconductors (e.g. aluminium, zinc and titanium oxides) show electron conduction, i.e. are n-type. Impurities may determine whether a semiconductor is n- or p-type, e.g. in germanium, silicon, lead sulphide, indium antimonide, etc.

Thermal equilibrium of the electrons in a semiconductor

If the valence electrons are but weakly bound to the atoms, thermal agitation will cause them to pass continually between the free and bound states. These states are not of the type implied by the usual meanings of free and bound.

Fermi statistics allow each cell of volume h^3/V in momentum space to contain not more than two electrons with opposite spins. An electron is bound if all the cells in phase space around it are filled with pairs of electrons; it cannot change its speed and coordinates. It can, it is true, pass to some remote free cell, but this would demand much extra energy, because all the nearby cells of smallest energy are full. Conversely, if the state is one in which many of the nearby phase cells are empty, the electron may pass from one cell to another, may change its energy or momentum, and so on. Such an electron is, for our purpose, free.

Thermal motion can transfer an electron in a semiconductor to a state of higher energy, which lies amongst the free phase cells. Sooner or later such an electron, which moves randomly with thermal motion or directionally because an external field is acting, meets a hole, an impurity atom, or a lattice defect, and becomes bound again because it loses potential energy. (The same is true of a hole.) Hence a free electron has a finite lifetime, which differs from electron to electron, as does the speed v. The lifetimes of the individual electrons show a statistical spread about some mean value τ_0.

An electron has time to make many collisions, and to follow a winding path, during this time τ_0. An electric field $\mathscr{E}$ acting during this time imposes on the random motion a displacement at a speed Δv, which is proportional to the field strength

$$\Delta v = u\mathscr{E}.$$

Here u is the mobility, or the mean rate of displacement caused by a

unit field. If there is no such field, the electron on an average moves randomly to a distance L away from its initial position within a time τ_0; this distance is called the diffusion length. The more frequent the collisions the smaller L and u. The path taken during τ_0 can be considered as being composed of separate straight sections, whose mean length l is called the mean free path. The speed built up by the field $\mathscr{E}$ in a length l determines u, which, from equation (15), can be put as

$$u = \frac{el}{mv}$$

where v is the mean speed of thermal motion; in this formula allowance is made for the statistical spread in the velocities.

If a semiconductor has a non-degenerate electron system

$$\tfrac{1}{2}mv^2 = \tfrac{3}{2}kT,$$

whence

$$u = \frac{el}{\sqrt{(3mkT)}}. \tag{41}$$

Einstein's equation relates the diffusion coefficient D to u

$$\frac{D}{u} = \frac{kT}{e},$$

whence

$$D = \sqrt{\left(\frac{kT}{3m}\right)}\,l. \tag{42}$$

We now relate L, l and τ_0 by assuming that the electron undergoes Brownian motion. In time τ_0 an electron of speed v covers a path of length $v\tau_0$. This path is made up of $v\tau_0/l$ sections of mean length l, and so the mean distance from the start of the path is

$$L = \frac{l}{\sqrt{3}}\sqrt{\left(\frac{v\tau_0}{l}\right)} = \sqrt{\left(\frac{v\tau_0 l}{3}\right)}. \tag{43}$$

We can also express D in terms of l and v:

$$D = \frac{lv}{3}.$$

We substitute for vl in equation (43) and get

$$L = \sqrt{(\tau_0 D)}.$$

Now u, D, l, L and τ_0 vary very much from one semiconductor to another; they depend largely on the presence of lattice defects and impurities. The purest and most carefully annealed germanium has, at room temperature, τ_0 equal to about 1000 μsec, L about 1 mm, u about 3600 cm^2/V sec, l about 10^{-5} cm and D about 0·1 cm^2/sec.

Temperature-dependence of the electron concentration and electrical conductivity

An equilibrium distribution over the available states and a steady electron concentration are set up, because some electrons are freed and others trapped. The higher the energy of any state the less often are electrons found in it. If the energy in the normal state is less by ΔE_0 than the energy in the free state, the free electron concentration n is given by a formula like that for electrolytic dissociation

$$n_0 = C \exp\left(-\frac{\Delta E_0}{2kT}\right). \tag{44}$$

Quantum theory (see Chapter 4) gives us

$$C = 2\frac{(2\pi mkT)^{\frac{3}{2}}}{h^3} \tag{45}$$

where the number of electrons equals the number of holes.

The mobility usually falls as the temperature rises. The type of lattice (valence, ionic, molecular) controls the manner in which u varies with T; a general formula is

$$u = AT^{-p}$$

where p may take a number of values; $p = \frac{3}{2}$ for atomic lattices.

The specific conductivity is

$$\sigma = nue.$$

We extract the temperature-dependent factors and get

$$\sigma = BT^{(\frac{3}{2}-p)} \exp\left(-\frac{\Delta E_0}{2kT}\right).$$

The second factor varies very much more rapidly with T than does the first if $\Delta E_0 \gg kT$, particularly since $T^{(\frac{3}{2}-p)}$ equals zero for atomic lattices. Hence in such cases we can assume that

$$\sigma = \sigma_0 \exp\left(-\frac{\Delta E_0}{2kT}\right) \tag{46}$$

without any large error. Taking logarithms, we have

$$\ln \sigma = \ln \sigma_0 - \frac{\Delta E_0}{2kT}. \tag{47}$$

Here $\ln \sigma$ is a linear function of $1/T$, as for ionic conductors. Fig. 25 shows $\ln \sigma$ against $1/T$, which is a straight line, whose slope is $\Delta E_0/2k$, which gives us ΔE_0. This is one of the most convenient ways of finding the ionization energy of the carriers.

In equation (46) σ_0 denotes σ at an infinite temperature when all the valence electrons are free. We have neglected the fact that u varies with T, because we supposed $T^{(\frac{1}{2}-p)}$ to be constant; we estimate σ_0 by taking u equal to $10^2\,\mathrm{cm^2/V\,sec}$. We then find σ_0 equals about $400\,\mathrm{ohm^{-1}\,cm^{-1}}$. It is not permissible, however, to extrapolate equation (46) to very high temperatures; in particular, ΔE_0 is rarely temperature-independent. It is found, though, that equation (46) is obeyed quite well at the temperatures at which semiconductors are usually used. We can take σ_0 as constant, but its value is seldom close to $400\,\mathrm{ohm^{-1}\,cm^{-1}}$, because u may be very different from our figure of $100\,\mathrm{cm^2/V\,sec}$; selenium has a mobility of about $1\,\mathrm{cm^2/V\,sec}$, for example, while u exceeds $3000\,\mathrm{cm^2/V\,sec}$ for germanium at 300° K, and may even be as high as $10^5\,\mathrm{cm^2/V\,sec}$ at low temperatures. The mobility of holes in a material is usually less than that of electrons.

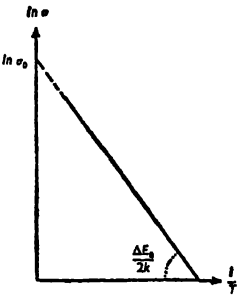


Fig. 25. Relation of electrical conductivity to temperature for a pure semiconductor, in $\log\sigma - 1/T$ coordinates.

The effects of impurities on the conduction mechanism and electrical conductivity

The electrical parameters of a semiconductor are very sensitive to the slightest changes in chemical composition. Excess atoms of one component, for example, act as sources of free charges (as in ionic conductors), whose dissociation energy E is often much less than ΔE_0. Even if the number of impurity atoms N is a minute proportion of the whole, the number of electrons which these atoms yield is very much larger than the number given by the lattice, especially at low temperatures.

The free carrier concentration given by impurity atoms whose valence electrons have an energy less by ΔE than their energy in the free state is

$$n_- = C_1 \exp\left(-\frac{\Delta E}{2kT}\right).$$

The valence band, also, always contains a certain number of holes.

The electron and hole concentrations are related by

$$n_- n_+ = n_0^2,$$

where $n_0 = C\exp(-\Delta E_0/2kT)$ is the concentration of either electrons or holes when no impurities are present.

The product of the two concentrations does not, in any given case, depend on the number of impurity atoms, or on the electron energies of such atoms. It always equals the square of the free electron or hole concentration for the pure semiconductor at that temperature.

If the impurity atoms are electronegative, and need a small energy $\Delta E < \Delta E_0$ in order to bind valence electrons, the hole concentration will be

$$n_+ = C_2 \exp\left(-\frac{\Delta E}{2kT}\right), \tag{48}$$

and the electron concentration will be

$$n_- = \frac{n_0^2}{n_+}.$$

If both electropositive and electronegative impurity atoms are included, the electrons given by the former will be taken up by the latter, so that positive and negative ions are formed respectively.† Let us suppose that 1 cm^3 of semiconductor contains N_1 electropositive impurity levels and N_2 electronegative ones, with $N_1 > N_2$. Then we get $2N_2$ ions of opposite signs, while the residual $N_1 - N_2$ electropositive levels make the substance an n-type semiconductor.

The number of electrons or holes of course cannot be larger than the number N of the impurity levels. Hence equation (48) is no longer true when the impurity electrons are exhausted. The free carrier concentration resulting from the impurities no longer rises with temperature, but an electron can be transferred from its normal bound state to a free state at all temperatures if an energy ΔE_0 is supplied; the corresponding concentration is given by equation (44). Free electrons arise from both sources only while the impurity levels are not exhausted. The total free electron (or hole) concentration is given by the sum of the two contributions, and the electrical conductivity is given by

$$\sigma = \sigma_1 \exp\left(-\frac{\Delta E_0}{2kT}\right) + \sigma_2 \exp\left(-\frac{\Delta E}{2kT}\right). \tag{49}$$

Both terms in equation (12) can be represented by straight lines if the formula is plotted as in Fig. 26. Line I represents the first term,

† In covalent semiconductors electropositive impurities yield holes while electronegative impurities yield electrons. The above results apply for ionically-bonded materials.

and is steeper than line II, which represents the second term. The $\log \sigma_1$ and $\log \sigma_2$ values are the intercepts on the ordinate axis at $T = \infty$, i.e. at $1/T = 0$. Now $\sigma_1 \gg \sigma_2$, so the first term makes a much larger intercept than does the second. The sum of the two terms (curve III) is not a straight line, but the high-temperature section is virtually line I, and the low-temperature section is virtually line II. The two straight sections are joined by a curved region representing intermediate temperatures.

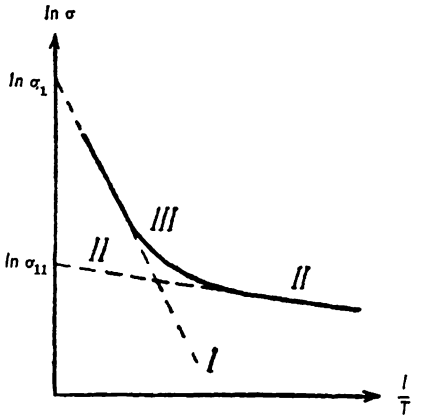


Fig. 26. Relation between electrical conductivity and temperature for an impurity semiconductor.

Zhuze and Kurchatov's[1] measurements, covering the range $-180°$ to 900° C for cuprous oxide, confirm this formula. Similar results have subsequently been given for silicon and for other materials (see, for example, Fig. 45). Now σ_2 increases with N, whereas ΔE falls only slowly while the impurity concentration is small (excess oxygen is the impurity in cuprous oxide); line II in Fig. 26 moves upwards, but stays at about the same slope. Fig. 27 shows the results of Zhuze and Kurchatov, which confirm this conclusion.

The impurities, or deviations from stoichiometric composition, determine the conductivity and the sign of the carriers. Excess metal atoms in *ionic* lattices supply electrons, whereas excess oxygen or sulphur (electronegative) atoms bind electrons, and cause hole conduction. The energy needed to release an electron from an electropositive atom, or to transfer an electron to an electronegative one, is usually much less than that needed to release an electron from the main lattice. Hence the impurities, if at all numerous, determine the magnitude and sign of the conductivity. Lead (or thallium) sulphide, if it contains excess metal, shows *n*-type conduction, and *p*-type conduction

if it has excess sulphur. PbS of near-stoichiometric composition has σ between 10^{-2} and 10^{-3} ohm^{-1} cm^{-1}; the value of σ becomes 100 to 1000 ohm^{-1} cm^{-1} if lead or sulphur is greatly in excess.[2]

Mixed electron and hole conduction is possible, but at room temperature is found only in low-resistance semiconductors, or in materials of exactly stoichiometric composition. It is rare in impurity semiconductors.

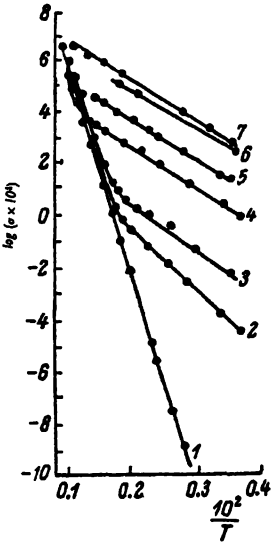


Fig. 27. Relation of electrical conductivity to temperature for cuprous oxide with various amounts of excess oxygen: (1) stoichiometric cuprous oxide; (2)–(7) successively increasing amounts of excess oxygen.

Neighbouring regions of a sample often differ greatly in conductivity; the cause is the large effect of even a trace of impurity, since 0·1% impurity can change the conductivity by a factor of a million. Traces of impurity of the appropriate valency change the conduction mechanism completely. Regions very close together may have thermal e.m.f.s of opposite signs in silicon, which is always to some extent impure.

Photoconduction

We have seen that the carriers are thermally activated in semiconductors, but not in metals. This means that the conductivity of a semiconductor tends to zero as the absolute zero of temperature is approached.

Sources other than thermal energy can free the charges, e.g. fast electrons, α-rays, light, and X rays. A bound electron gets an amount of energy $h\nu$ by absorbing a light quantum of frequency ν; h equal to

$6{\cdot}62 \times 10^{-27}$ erg sec is Planck's constant. The electron passes to the free state if $h\nu > \Delta E_0$ and so assists in carrying a current. Such photoelectrons have the same mobility as do electrons of thermal origin. Zhuze and Ryvkin,[3] and also Putseiko,[4] have given many data that show that the photocarriers usually have the same sign as the thermal carriers.

The extra conductivity produced by these electrons (or holes) is termed the photoconductivity; only quanta that satisfy the condition

$$h\nu \geqslant \Delta E_0 \tag{50}$$

can produce photoconduction.

An electron so released soon returns to a vacated state, or becomes trapped by an impurity atom. An equilibrium is reached between the rates of production and loss of photoelectrons. This differs from the thermal equilibrium in that the concentration of free electrons (or holes) is increased, and, thus, the conductivity is also increased, because the mobility is unchanged. Hall effect measurements have shown that photoconduction in cuprous oxide is caused almost entirely by holes.

If impurities are absent, and the thermal carriers few, the number n_1 of carriers freed per sec by light of quantum energy $h\nu$ is proportional to the light flux L

$$n_1 = k_1 L, \tag{51}$$

while the number of charges lost by recombination per sec is n_2; these numbers are equal at equilibrium. Now n_2 is proportional to the number of free electrons n, and to the number of holes left behind (also n)

$$n_2 = k_2 n^2,$$

and so

$$n = \sqrt{\left(\frac{k_1}{k_2}\right)} L^{\frac{1}{2}}. \tag{51a}$$

We have neglected the thermal electrons, because we have assumed that they are few in number relative to the photoelectrons. We now assume that thermal agitation produces n_0 electrons per sec. Then equation (51) is replaced by

$$n_1 = k_1 L + n_0. \tag{52}$$

The conditions for the reverse transitions are the same for both types of electron, so the equilibrium condition is

$$n_0 + k_1 L = k_2 n^2$$

and hence

$$n = \frac{1}{\sqrt{k_2}} \sqrt{(k_1 L + n_0)}. \tag{52a}$$

If the light produces relatively few electrons, i.e. if $n_0 \gg k_1 L$, we may expand the root in series form and take only the linear term

$$n = \sqrt{\left(\frac{n_0}{k_2}\right)}\left(1+\frac{1}{2}\frac{k_1 L}{n_0}+\ldots\right),$$

because in the dark the free electron concentration n' is given by

$$n_0 = k_2 n'^2,$$

i.e.

$$n' = \sqrt{\left(\frac{n_0}{k_2}\right)}.$$

The extra number of carriers is

$$n-n' = \frac{1}{2}\sqrt{\left(\frac{n_0}{k_2}\right)}.\frac{k_1 L}{n_0} = \frac{1}{2}\sqrt{\left(\frac{k_1^2}{k_2 n_0}\right)}L, \tag{53}$$

so the additional conductivity is proportional to the energy absorbed. If, on the other hand, $n_0 \ll k_1 L$, we have from equation (51a) that

$$n = \sqrt{\left(\frac{k_1}{k_2}\right)}\sqrt{L}.$$

Now n_0 rises rapidly with T so that the photoconductivity falls, not only relative to the dark conductivity

$$\frac{n-n'}{n} = \frac{1}{2}\sqrt{\left(\frac{k_1}{n_0}\right)}L,$$

but also in absolute value

$$n-n' = \frac{k_1}{2\sqrt{(k_2 n_0)}}L.$$

If N, the number of impurity atoms, is large, we have

$$n_1 = n_0 + k_1 L \quad \text{and} \quad n_2 = k_2 N n,$$

so

$$n = \frac{n_0 + k_1 L}{k_2 N}. \tag{54}$$

The number n' at equilibrium in the dark is given by

$$n_0 = k_2 N n',$$

i.e.

$$n' = \frac{n_0}{k_2 N}.$$

The extra number of carriers is

$$n-n' = \frac{k_1}{k_2 N}L. \tag{55}$$

Hence the photoconductivity is proportional to the light intensity if the light is weak or if the dark conductivity is high, and is proportional

to the square root of the intensity if the light is strong or the dark conductivity is low.

Arsen'eva and I[5] and Gyulai[6] have shown that the photoconductivity is directly proportional to the number of photons absorbed for coloured rock salt; the same has been found to be true for semiconductors in which the impurity atoms yield the photoelectrons.

A. V. Ioffe and I[7] have shown that light which is absorbed strongly may reduce the conductivity, instead of increasing it (the negative photoelectric effect). Here the photoelectrons diffuse into nearby regions and accelerate recombination (reduce the carrier concentration). Cases are also known in which the photoelectrons increase the free carrier concentration by secondary processes to such an extent that the total extra concentration is much larger than the concentration of the photoelectrons alone.

The conductivity may be very greatly increased by but a few photoelectrons if the material consists of small crystals separated by insulating layers.

An absorbed photon does not necessarily produce a photoelectron; there are many opaque crystals that show no photoconductivity. The light energy is transformed to thermal energy in other ways. There are reasons for supposing that the primary effect is the production of an electron–hole pair, which diffuses as a whole. Frenkel' has predicted this effect from theory, and has called the combination an exciton.[8] No photoconductivity is found if the charges simply recombine, and give up their energy to the lattice, but luminescence may then occur; if impurities or thermal fluctuations cause the pair to separate, and one is trapped, the other can give a photocurrent. Zhuze and Ryvkin[9] have shown that this process occurs in cuprous oxide. The light is absorbed by the lattice, but the photoconductivity is determined by the impurity content, and by ΔE (not by ΔE_0).

The photocurrent can be described in terms of the equilibrium electron concentration n and the mobility u, as for the dark current. Photoconduction can be described more physically in another way if the dark current is negligible (as in insulating crystals); this description is also applicable to semiconductors.

We denote the amount of light striking the surface per sec by L and the absorption coefficient by x; the number of electrons freed per sec is then

$$n_1 = \frac{xL}{h\nu}. \tag{56}$$

If only some fraction α of the photons release electrons, we have

$$n_1 = \alpha \frac{xL}{h\nu}. \tag{56a}$$

Any one of these electrons is available to carry current for a time τ; τ varies about some mean value τ_0. An electric field of strength $\mathscr{E}$ displaces the electrons on an average by a distance d equal to $u\mathscr{E}\tau_0$ in its own direction. If this field is produced by two parallel electrodes a distance D apart (Fig. 28) and having a potential difference V, then

$$d = u\frac{V}{D}\tau_0.$$

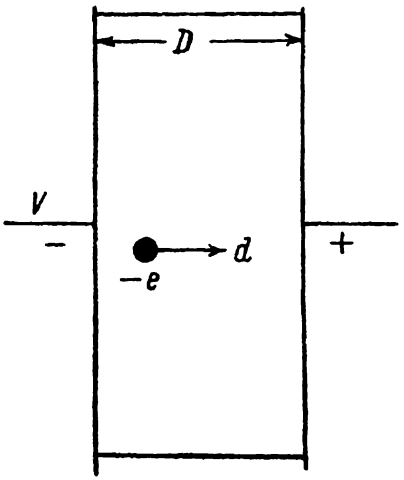


Fig. 28. Experiment to measure the mean lifetime of a photoelectron.

An electron displaced by a distance d within this plate condenser induces a negative charge q on the positive electrode, where

$$q = e\frac{d}{D}.$$

A similar positive charge is produced on the negative electrode. The light frees n_1 electrons per sec per cm^3, so $n_1 SD$ electrons are freed per sec in an area S between electrodes a distance D apart. The current that flows in a wire joining the electrodes is

$$J = n_1 eu\frac{V}{D}S\tau_0. \tag{57}$$

This current can be expressed in terms of the equilibrium electron concentration as

$$J = neu\frac{V}{D}S. \tag{58}$$

The two expressions, combined, give

$$n = n_1 \tau_0. \tag{59}$$

Now equation (56) gives us n_1, and the Hall effect gives us n, so τ_0 can be found; the values range from 10^{-8} to 10^{-7} sec. The mean thermal speed at room temperature is 10^7 cm/sec, and the mean free path is about 10^{-7} cm, so an electron can make at least 10^7 collisions while it is free. This number is so large that we may say that the electron has the equilibrium thermal energy for a large part of its lifetime, even though its initial energy is high, because each collision can cause the electron to exchange energy with the lattice. This is confirmed by the fact that the photocarriers and thermal carriers have the same mobility even in crystals in which the mean free path depends on energy (e.g. ionic crystals).

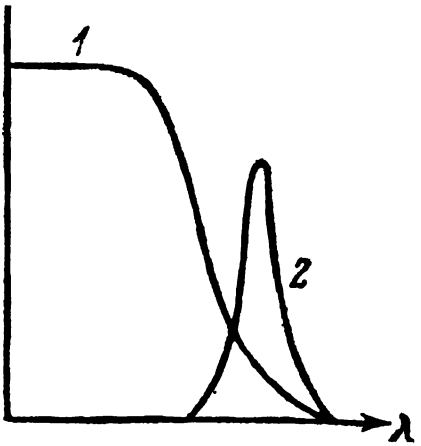


Fig. 29. (1) Optical absorption spectrum, and (2) photoconductivity spectrum for a crystal.

The condition $h\nu > \Delta E$ is necessary, but not sufficient, for photoconduction to occur. Pohl and Gudden[10] have given some empirical rules, namely that (1) photoconduction unrelated to impurities or to lattice defects occurs in all insulators and semiconductors that have refractive indices of over 2 (i.e. dielectric constants >4), and that (2) photoconduction occurs only at the long-wave edge of the absorption spectrum. Fig. 29 illustrates the relation between the absorption and photoconduction spectra.

The cause of the sharp drop in photosensitivity in the region where the absorption is strong has not yet been discovered. It may be that, because the light is all absorbed in a layer less than a micron thick, the photoelectron concentration is high, and recombination is accelerated;

the result is a drop in the free carrier concentration. Ryvkin has shown that no increased conductivity is found if some other means (e.g. fast electrons) is used to produce electrons or holes in such thin layers.[11]

The energy ΔE_p required for photoionization may differ from the energy ΔE_t needed for thermal ionization for the following reasons:

1. Ionization events caused by light or X-rays obey selection rules, whereas those caused thermally do not; the energy $h\nu$ must be sufficient, and the momentum $h\nu/c$ must be appropriate. This means that an electron in an s-state (which has an orbital moment $l = 0$) can be transferred by light only to a p-state (with $l = 1$); one in a p-state can go either to an s-state, or to a d-state ($l = 2$), etc. The selection rules do not apply to thermal transitions. A phonon of frequency ν_3 (energy $h\nu_3$) has a momentum of $h\nu_3/v_3$, where v_3 is the speed of sound, which is 10^5 times smaller than the speed of light; hence a phonon has 10^5 times the momentum of a photon with the same energy. All transitions are then equally probable.

TABLE 5

	Bond type	ΔE_p (eV)	ΔE_t (eV)
Boron	Valence	0·96–1·27	0·88–1·28
Diamond	,,	5·3	5·2±0·3
Silicon	,,	1·15	1·12
Germanium	,,	0·73	0·72
Phosphorus	Molecular	1·4–1·6	1·5–1·6
Arsenic	,,	1·2	1·2
Sulphur	,,	2·4	2·6
Selenium	,,	1·5–1·9	~1·7
Tellurium	,,	0·37	0·34
Iodine	,,	1·30	1·24
Cuprous oxide	Partly ionic	1·95	1·45–1·84
Lead telluride	,,	0·62	0·44–0·60
Zinc oxide	,,	3·0	2·42
Titanium oxide	,,	3·2	3·05

The minimum energy needed to reach the conduction band, ΔE_t, may correspond to a forbidden optical transition; an energy $\Delta E_p > \Delta E_t$ is needed in order to satisfy the selection rules.

2. ΔE_p and ΔE_t may differ in ionic or dipolar crystals because electron transitions involve a shift in the ions or dipoles. Such shifts occur very slowly relative to electron transitions; the respective times are about 10^{-12} and 10^{-15} sec. The electrons alone have time to move during optical transitions; the ions follow the charges and take up

their equilibrium positions (i.e. positions of lower potential energy) much later. The energy they release appears as thermal energy.

The ions do have time to come to equilibrium during thermal transitions; then the electrons have lower final energies so that the total energy needed, ΔE_t, is less than that for optical transitions.

The temperature variation of σ (the slope of the $\log \sigma$ against $1/T$ plot at high temperatures) can give us ΔE_t. ΔE_p is given by the long-wave limit of the photoconductivity spectrum, i.e. by the minimum frequency capable of giving the internal photoelectric effect, or by the long-wave limit to the absorption band.

It is found, however, that ΔE_p and ΔE_t differ little for all crystals with valence bonds, and for many crystals with ionic bonds (Table 5).

Volume photo-e.m.f.s

The electron and hole concentrations vary from point to point if a semiconductor is unevenly illuminated. The free carriers diffuse from regions which are brightly illuminated to regions where the light is not so strong, and so cause electric fields. These fields oppose the motion of charges of one sign, and accelerate charges of the other sign. Potential differences as large as 0·1–0·2 V may be set up; e.m.f.s of this order have in fact been found with single crystals of cuprous oxide, antimony sulphide, etc. A. V. Ioffe and I[12], Lashkarev,[13] and Lyashchenko have studied these effects; Landau[14] has given a theory of them. These e.m.f.s occur only if electrons and holes both take part in the diffusion. It is easy to see the reason for this. If carriers of only one sign are produced, they leave behind them charges of opposite sign, which soon prevent further diffusion, whereas if charges of both signs diffuse away there is no space charge left behind; thus, diffusion can continue until the electric field makes the rates of diffusion of electrons and holes equal.

Kikoin and Noskov[15] have detected a new effect, namely, that e.m.f.s of up to 15 V are produced by illuminating cuprous oxide cooled to liquid-air temperature in a magnetic field. The magnetic field can act only on moving charges, and so it must be that the currents of electrons and holes set up by the uneven illumination are deflected. Kikoin and Ansel'm have given a theory of the effect which agrees exactly with experiment.[16] The Kikoin–Noskov effect is still more marked at helium temperatures.

Strong electric fields and breakdown

Ohm's law applies to a semiconductor if the extra velocity Δv gained by an electron within one mean free path is negligible relative to its thermal velocity v. Hence the duration τ of the free path is independent

of Δv, which is produced by the field $\mathscr{E}$ in time τ

$$\Delta v = \frac{e\mathscr{E}}{m}\tau$$

(where τ is independent of $\mathscr{E}$); Δv is proportional to $\mathscr{E}$, and so

$$u = \frac{\Delta v}{\mathscr{E}} = \frac{e}{m}\tau.$$

u is also independent of $\mathscr{E}$.

The current density j obeys Ohm's law, i.e. is proportional to $\mathscr{E}$

$$j = neu\mathscr{E} = \sigma\mathscr{E}.$$

Ohm's law requires that nu and σ are independent of $\mathscr{E}$. This indicates the limits of its applicability. Ohm's law does not apply if $\mathscr{E}$ is so large as to alter n, or so large that Δv becomes comparable with v. The second case occurs at about 1000 V/cm, whereas n is not affected until $\mathscr{E} > 10{,}000$ V/cm.

Let us consider the second condition first. We assume $\mathscr{E} =$ 1000 V/cm, the free path $l = 10^{-6}$ cm, and $3kT/2 = 0{\cdot}04$ eV (which implies that $v = 1{\cdot}1 \times 10^7$ cm/sec), and $\tau = 0{\cdot}9 \times 10^{-13}$ sec, which gives $\Delta v = 1{\cdot}6 \times 10^5$ cm/sec, i.e. 1·5% of v; $\mathscr{E} = 10^4$ V/cm gives $\Delta v = 0{\cdot}15v$.

The deviation from Ohm's law is a function of the lattice type. Diamond, silicon, germanium and similar lattices (pure valence) have l independent of v; τ is reduced proportionally when v is increased by Δv, and the collisions become more frequent. On an average τ, and hence u, is reduced. It is found that the deviations from Ohm's law become appreciable at room temperature when $\mathscr{E}$ is 900 V/cm for electron conduction or 1400 V/cm for hole conduction in germanium; the corresponding figures for silicon are 2500 V/cm and 7500 V/cm. Landau's explanation,[17] which agrees well with experiment, shows that the current density is proportional to $\mathscr{E}^{\frac{1}{2}}$, and not to $\mathscr{E}$, in strong fields.

Other conditions occur in ionic lattices, in which the mean free paths increase with v. We then expect that a rise from v to $(v+\Delta v)$ will increase τ, and hence u also. Rekalova has found that Sb_2S_3 has a mobility that increases exponentially with $\mathscr{E}$ for fields of about 10^6 V/cm.[18]

The mobility may increase with the field also on account of the mechanism by which carriers are scattered by ionized impurity atoms. The scattering probability and scattering angle both grow less as v increases. Hence l increases with v, as in ionic crystals. This effect is found in germanium at 20° K when the field is of the order of 10 V/cm;

the scattering is then caused mainly by impurity ions, and l is very large (about 10^{-4} cm).

There is a field $\mathscr{E}_0$ (about 10^5 V/cm) at which the conductivity starts to rise rapidly in accordance with Poole's law (Fig. 30)

$$\sigma = \sigma_0 \exp[\alpha(\mathscr{E} - \mathscr{E}_0)].$$

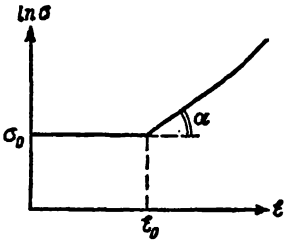


Fig. 30. Relation of conductivity σ to field strength $\mathscr{E}$. $\mathscr{E}_0$ is the critical field; $\sigma = \sigma_0 =$ constant when $\mathscr{E} < \mathscr{E}_0$, and $\sigma = \sigma_0 \exp[-(\mathscr{E} - \mathscr{E}_0)]$ when $\mathscr{E} > \mathscr{E}_0$.

We have examined this effect for many semiconductors, and have found that it is caused by an increase in n, and not by an increase in u.[19] One experiment that led to this conclusion was based on the assumption that Δn, the number of photoelectrons, is independent of the field. The assumption is not, strictly speaking, justified, because the field may affect the recombination probability, and hence the equilibrium photoelectron concentration. We may, however, assume that Δn depends little on $\mathscr{E}$, and so we can calculate the mobility from the rise ΔJ in the current

$$\Delta J = \Delta n e u \mathscr{E}$$

whence

$$u = \frac{\Delta J}{\mathscr{E} e \Delta n}.$$

We subtract the dark current J_1 from the total current J_p found when the semiconductor is illuminated, and thus get ΔJ; we find that σ_l equal to $J_1/\mathscr{E}$ rises very rapidly with $\mathscr{E}$, whereas $(\sigma_p - \sigma_l) = \Delta\sigma = J/\mathscr{E}$ is either constant (Fig. 31, a) or else rises very much more slowly than does σ_l (Fig. 31, b). We have no reason to suppose that the photoelectrons differ in mobility from the thermal electrons. Engelhardt has shown that the two mobilities are the same for cuprous oxide. These results thus show that σ rises rapidly with $\mathscr{E}$ because n increases, and not because u changes.

Another test, carried out with a semiconductor wafer, also confirmed this conclusion. A strong field $\mathscr{E}$ was applied across the wafer, and a weak one $\mathscr{E}_0$ along its length; the mobility in the latter field was certainly constant. However, the strong field $\mathscr{E}$ caused the lengthwise current to increase just as rapidly as did σ in the direction of $\mathscr{E}$. Hence the rise in σ is caused by the rise in n, and not by any change in u.

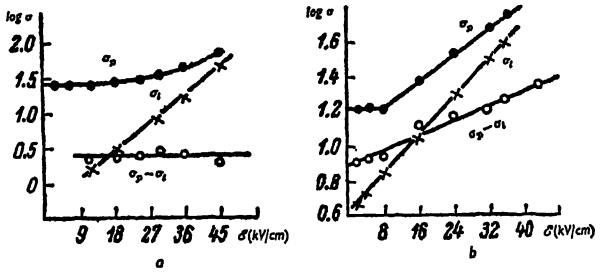


Fig. 31. Dependence of dark conductivity (σ_t) and photoconductivity (σ_p) on electric field strength.

A rise in conductivity does not of itself cause the semiconductor to break down electrically, because the current may be too small to cause damage. Breakdown is accompanied by a marked instability in the current, which shows that some new process is operating. There have been many papers on the breakdown mechanism, but no final conclusion has yet been reached. It may be that breakdown occurs when the free electrons gain more energy from the field than they can transfer to the lattice. The rapidly rising kinetic energy causes the number of free charges to rise, and the specimen breaks down. There is also a thermal breakdown mechanism, when the raised conductivity causes heating, which itself raises the conductivity further, and so on. The catastrophic rise in temperature finally melts or evaporates the material at the point where the current density is highest.

The barrier layer

A peculiar effect (a barrier layer) is found at a contact between a metal and a semiconductor. The effect is much used in technical applications.

The outermost layer of a semiconductor sometimes differs in composition from the inner parts. Loss of oxygen or sulphur, or absorption of atmospheric oxygen, may cause a surface layer to be much altered, or to become non-conducting. A simple example is the layer of oxide on aluminium metal, or the layer of cuprous oxide at an interface with copper, which is readily oxidized.

Even if such chemical changes are absent, we find layers at junctions with metal electrodes, or with other semiconductors, if the contact potentials are very different, which layers display a sharp change in conductivity.

The difference in contact potentials indicates that the binding energies for electrons are different. The electrons (or holes) pass through any such contact from the side where their energy is high to the side where it is lower until the sides have acquired charges sufficient to balance out the difference in contact potential. Metals can provide such charges from layers of a thickness d equal to 10^{-8} cm (one atom thick), because their conductivities are high. Semiconductors need much thicker layers; the lower the conductivity the thicker the layer. We find that d is about 10^{-5} cm if σ is 10^{-3} ohm^{-1} cm^{-1}, and about 10^{-3} cm if σ is 10^{-7} ohm^{-1} cm^{-1}. The space charge present in this layer increases or decreases the carrier concentrations in accordance with its sign. Impurity ions remaining after the electrons have left give a positive space charge in n-type conductors; the layer becomes a very much worse conductor. If the holes are more mobile, on the other hand, the positive charge causes them to accumulate near the electrode in numbers more than sufficient to balance the negative ions, so the conductivity is increased. Similar layers form at a junction between two semiconductors with different contact potentials.

One result of this is that the conductivity of the semiconductor depends very much on the contact potential of the electrodes. I have shown that n-type conductors such as TiO_2, WO_3 and ZnO have very much increased resistivities if electrodes with higher potentials (e.g. gold, graphite) are used.[20] On the other hand, p-type conductors (Cu_2O, UO_2, V_2O_5) show higher resistivities with metals having low work functions, such as Mg and Al.

Almost all the potential drop across the conductor is concentrated in such a layer, if the resistivity there is very much higher than elsewhere; the layer is only about 10^{-5} cm thick, so the field can become tens of thousands of volts per cm even if the total potential drop is only a few volts. The conductivity increases rapidly with the field at such field strengths. Hence we can choose the electrodes so as to make a conductor whose resistance falls rapidly as the applied potential difference increases above about ten volts.

One body acquires a negative charge, and the other a positive one, when two bodies that differ in contact potential are brought together. This effect explains electrification by friction or contact, which has long been known, but which has not previously been explained.

This relation between the conductivity and the applied potential has important practical uses. The current across the boundary between

a p-type and an n-type conductor is largest when the holes and electrons move in opposite directions. If we apply an external potential difference to the insulating layer formed by removing the carriers from an n-type semiconductor so as to extract yet more carriers, the layer increases in thickness and resistance. A field in the reverse direction makes the layer thinner and more highly conducting.

The ratio of forward and reverse currents given by a contact between a metal and a semiconductor (the rectification coefficient) is, however, small compared with the values found in technical rectifiers, which may be many thousand. A. V. Ioffe has shown that these large values are found only at junctions between semiconductors with hole and electron conduction mechanisms (p–n junctions). It has been found that such junctions can occur even within a single crystal. The resistance in the forward direction is very low, and in the reverse direction very high. An alternating voltage applied to such a device gives a large current in one direction only; the alternating current is rectified. Semiconductor rectifiers are now extensively used in technology.

Barrier layers can also be used in producing a new type of photocell. If a photosensitive semiconductor has a surface barrier layer, the photoelectrons can pass through the barrier and charge a transparent electrode on top of the barrier layer; a potential difference which may be some tenths of a volt can be produced. This potential difference can cause a fairly large current to flow in an external circuit; in some cases (Tl_2S, Ag_2S, CdS) the values can rise to 10 mA/lumen, or with Ge to 50 mA/lumen, which are at least a hundred times better than the best vacuum photocells give. Barrier photocells need no external supply; the energy comes from the light absorbed. Silicon photobatteries have recently been made with conversion efficiencies of about 11%. Lashkarev and Lehovek have developed the theory of barrier photocells.[13, 21]

Ionizing radiations also can increase the free-carrier concentration, and so can produce e.m.f.s if barrier layers or p–n junctions are present. Much of the radiant energy (up to half) can in certain cases be turned into electrical energy.

The need for the quantum theory in studies on semiconductors

So far we have given only a purely phenomenological description of the more important effects found in semiconductors, with elementary explanations of the mechanisms. More detailed and exact explanations can be given only from quantum theory, which explains the experimental facts, and often agrees very well with them. The most important feature of semiconductors (hole and electron conduction), which causes the signs of the thermal e.m.f. and of the Hall effect to depend on the nature of the semiconductor, stems from quantum laws, and can be

understood only in terms of the quantum theory. This theory describes the laws of motion for electrons in a solid, and the way in which the electrons interact with the lattice.

We have so far used only one aspect of quantum mechanics, namely, Fermi statistics for electrons, but even here we have not dealt with the energy-level systems, or with the statistical weights of the various levels. We shall treat the data we have given as a first survey of the facts; we now turn to quantum ideas on the behaviour of electrons in a semiconductor, and will then pass on to describe the properties of semiconductors more fully. We shall not present the mathematical theory in full, nor shall we give rigorous proofs; only the main ideas and results are dealt with, since these are needed for us to survey the properties of semiconductors systematically, and to indicate the main methods of controlling these properties for technical purposes. Two monographs on the theory of semiconductors (by Ansel'm and Samoilovich) are to appear in the near future.

Much work has been done abroad on semiconductor theory, and much also by Landau and Pekar (polaron theory), by Frenkel' (excitons), by Davydov, Blokhintsev and Gubanov (rectification), by Adirovich (luminescence), and by Ansel'm (excitons); the work of the Americans Bardeen and Shockley has been especially important in the last few years. Many theoretical and experimental studies are in hand at present in the Soviet Union and abroad. It will be clear from what follows that there are still, however, many major theoretical problems unsolved, with the consequence that many important technical problems are solved by trial and error without theoretical guidance.

References

1. V. P. Zhuze and B. V. Kurchatov, "The electrical conductivity of cuprous oxide", *Phys. Zs. S.U.*, **2**, 453 (1932).
2. T. L. Koval'chik and Yu. P. Maslakovets, "The effects of impurities on the electrical properties of PbTe", *Zhur. Tekh. Fiz.*, **26**, 2417 (1956).
3. V. P. Zhuze and S. M. Ryvkin, "Determination of the sign of photocurrent carriers", *Dokl. Akad. Nauk SSSR*, **62**, 55 (1948).
4. E. K. Putseiko, "Use of the condenser method in the determination of the sign of photocurrent carriers", *Dokl. Akad. Nauk SSSR*, **67**, 1009 (1949).
5. A. Arsen'eva, "Photoconduction in rock salt", *Zs. f. Phys.*, **37**, 701 (1926).
6. Z. Gyulai, "Absorption in photoconducting NaCl crystals", *Zs. f. Phys.*, **33**, 251 (1925).
7. A. V. Ioffe and A. F. Ioffe, "Spectral distribution of the internal photo-effect in cuprous oxide", *Zhur. Eksper. Teor. Fiz.*, **6**, 737 (1936).
8. Ya. I. Frenkel', "On the mechanism of light absorption in solid bodies", *Phys. Rev.*, **37**, 17, 459, 1276 (1931).
9. V. P. Zhuze and S. M. Ryvkin, "Photoconduction in cuprous oxide", *Izv. Akad. Nauk SSSR, Ser. Fiz.*, **16**, 93 (1952).

10. B. Gudden, *Lichtelektrische Erscheinungen* (*Photoelectric Effects*). Springer, Berlin, 1928.
11. S. M. Ryvkin, B. M. Konovalenko and Yu. S. Smetannikova, "Dependence of the induced conductivity in foils and single crystals on the energy of excited electrons", *Zhur. Tekh. Fiz.*, **24**, 961 (1954).
12. A. V. Ioffe and A. F. Ioffe, "Photo-e.m.f.s in cuprite crystals", *Phys. Zs. S.U.*, **7**, 343 (1935).
13. V. E. Lashkarev, "Photo-e.m.f.s in semiconductors", *Izv. Akad. Nauk SSSR, Ser. Fiz.*, **16**, 18 (1952).
14. L. D. Landau, "On the theory of the photo-e.m.f. in semiconductors", *Phys. Zs. S.U.*, **9**, 477 (1936).
15. I. K. Kikoin and M. M. Noskov, "A new photomagnetic effect in cuprous oxide", *Zhur. Eksper. Teor. Fiz.*, **4**, 123 (1934).
16. A. I. Ansel'm, "Theory of the photomagnetic effect in semiconductors in strong magnetic fields", *Zhur. Tekh. Fiz.*, **24**, 2064 (1954).
17. L. D. Landau and A. Kompaneets, "Departure from Ohm's law in semiconductors in strong electric fields", *Phys. Zs. S.U.*, **6**, 163 (1934).
18. G. I. Rekalova, "Photoconductivity of As_2S_3 in strong electric fields", *Zhur. Tekh. Fiz.*, **21**, 1041 (1951).
19. A. V. Ioffe and A. F. Ioffe, "Semiconductors in strong electric fields", *J. Phys. U.S.S.R.*, **2**, 283 (1940).
20. A. V. Ioffe, "Contact resistances between semiconductors and metals", *Zhur. Eksper. Teor. Fiz.*, **15**, 721 (1945).
21. K. Lehovec, "Theory of photoeffects in barrier layers", I, *Zs. f. Naturforsch.*, **1a**, 528 (1946); II, *Ibid.*, **2a**, 398 (1947).

CHAPTER

4

Fundamentals of the Quantum Theory of Semiconductors

The quantum states of electrons in a solid

We have already brought in some ideas from quantum mechanics and quantum statistics in considering the way in which energy is spread among the electrons. We noted that the main aspects of the theory have been confirmed from the X-ray spectra, and that the contradictions in the classical theory have been removed. The simplified theory we used to replace the classical electron gas theory is, however, not adequate to explain many of the properties of metals and semiconductors. The state of an electron was specified in terms of its energy alone; the parameters specifying its state in a crystal lattice, which depend on direction in that lattice, were ignored. These factors are of paramount importance in semiconductors and cannot be neglected. A semiconductor does not consist of an isotropic volume filled with electrons, but is a set of atoms or molecules so spaced that the wave functions of electrons on adjacent atoms overlap.

We first deal with the properties of electrons in free atoms. Here the electrons fall into groups of states, denoted by K, L, M, etc., that are determined by the principal quantum number n, which can take the values $1, 2, 3, \ldots$. The second quantum number l takes the values $0, 1, 2, 3$, which differentiate the s, p, d and f states. The other two quantum numbers, j and m, define the orientations of the magnetic fields. The energies of electrons which differ only in their spin orientation are almost equal; to each energy there correspond two electrons with opposed spins.

Quantum theory shows that the s states can take two electrons (of opposed spins), the p states six, the d states ten, etc. The numbers of electrons actually present in any given state of the atom are shown by superscripts to the s, p, d and f symbols; the 11 electrons in the Na atom are specified by $1s^2 2s^2 2p^6 3s^1$, for instance. The last ($3s^1$) electron is the valence electron and is quite easily removed (5 eV is required). This $3s^1$ electron can enter several excited states (of higher energy)

without leaving the atom; the energies required range up to the ionization energy. After removing this electron one is left with the Na^+ ion, which has a distribution $1s^2 2s^2 2p^6$ resembling that of the neutral Ne atom, but with a larger nuclear charge (11 instead of 10).

We consider the transition from isolated atoms or molecules to the solid state as a process in which the particles are gradually brought together. We assume that the atoms are, at first, in the same relative positions as in the crystal but so far apart that their interactions may be neglected. The array is compressed uniformly; we thus reach the crystalline state, where the properties depend on direction.

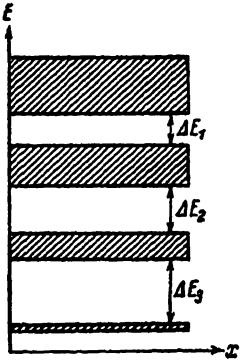


Fig. 32. The energy bands in a one-dimensional solid. $\Delta E_{1,2,3}$ are the distances between the allowed bands.

A large number N of identical atoms, brought together to produce a solid or liquid, interact in such a way that every quantum state of the isolated atom is split into N distinct states. The discrete levels of the free atom go over to a system of bands, each of which consists of N levels very close together (Fig. 32); the levels are spaced at about 10^{-22} eV, so the whole set can be taken as forming a continuous band about 1 eV wide. We call such a system of levels a band, though it has, in fact, just N and only N energy levels.

The band of levels in the solid will be only partly filled if the atoms have incomplete valence electron groups, e.g. as with the one electron in the $3s$ state in sodium. A full s state contains two electrons. This state is split into N levels in the solid or liquid; each level can take two electrons differing in spin direction only. The $3s$ band can thus take $2N$ electrons; the $N/2$ lowest levels have two electrons each, while the top $N/2$ levels are empty.

Now consider a crystal of rock salt (NaCl), which is an insulator. The $3s$ electron of sodium is coupled to the five $3p$ electrons of chlorine

(the state of which can be represented as $1s^2\,2s^2\,2p^6\,3s^2\,3p^5$), so yielding a Cl^- ion with a closed shell, $1s^2\,2s^2\,2p^6\,3s^2\,3p^6$, that resembles the neutral atom of the inert gas argon. The Na^+ and Cl^- are bonded mainly by their electrostatic attraction. The numbers of states in the Na^+ and Cl^- ions correspond exactly to the numbers of electrons. The $2p$ band of Na and the $3p$ of Cl are completely filled. The $3s$ levels of Na, and all the excited states of Cl, are completely empty.

These distributions may be compared with those of the isolated atoms; in Na, as in sodium metal, some of the $3s$ levels would stay free. In NaCl the Na^+ ions are much further apart than are the Na atoms in the metal; the $3s$ wave functions thus overlap very much less, and the $3s$ levels of Na lie some 6 eV above the $3p$ levels of Cl (this is why the $3s$ electrons pass over to the chlorine $3p$ levels). The $3s$ band in sodium metal is only partly filled, but in NaCl it is completely empty; the electron energy is some 6 eV higher than in the $3p$ band, so the energy is least if the $3p$ band of the Cl atoms is entirely filled, while the $3s$ band of sodium stays quite empty.

The electrons in these full bands behave quite differently from the electrons in metals which we dealt with previously. Any force on an electron tends to cause its state to change, i.e. its speed or direction of motion to change, its spin to reverse, or its energy to be altered. However, all the otherwise accessible states are full (since each state can accommodate one electron only), so no changes are possible; in particular no changes in the state of motion (currents) are possible.

Free and bound, as applied to electrons in an electronic semiconductor, have meanings different from those which they have as applied to ions. A bound ion is one held by chemical or electrostatic forces at a potential minimum, which is so deep that the work needed for its escape exceeds kT greatly. If the energy does not change on displacement, or the minimum is so shallow that thermal motion readily moves the ion from one minimum to another, the ion is free and an external field can affect its motion.

An electron in a solid is free if the energy structure has enough states, with separations which are small relative to kT, that are not taken up by other electrons. External forces or thermal motion may change the state of such an electron by transferring it to a state of similar energy. An electron is bound if the nearby states it would otherwise enter are already taken up. Bound electrons carry no current, but this does not mean that an external field has no effect; the quantum states are altered. The equilibrium positions are moved somewhat but stay on the same atoms; the shifts are small and are proportional to the field $\mathscr{E}$. This dielectric displacement lags behind the field by not more than 10^{-14} to 10^{-15} sec, and weakens $\mathscr{E}$ by a factor ϵ_0, called the high-frequency

dielectric constant of the medium. This constant equals the square of the refractive index for light of the corresponding frequency.

Any ions or dipoles are also moved by the field and cause an extra dielectric polarization ϵ', which takes much longer (10^{-11} to 10^{-12} sec) to reach equilibrium. Thus the dielectric constant ϵ has two components

$$\epsilon = \epsilon_0 + \epsilon'.$$

The possible states for the electrons are not limited to those in the full bands that result from the splitting of the ground states; excited and ionized states also give bands. These states have more extended wave functions than have the ground states; thus there is much greater splitting and the bands overlap. The bands given by the excited and ionized states fuse into one free electron or conduction band. Isolated atoms have ionization energies of 5 to 20 eV; the free states in a solid are only one or, at most, a few eV away from the ground states.

The bands from different atomic states may overlap to give one common band in the solid. The total number of levels equals the sum of the numbers for the separate bands, while the intermediate forbidden bands vanish. However, the differences between the various types of state persist.

All the electrical interactions are weakened if the dielectric constant is high; the discrete atomic levels go over to broad bands. The forbidden band always has a width much less than the distance between the ground and excited states.

The rules for isolated atoms still apply to solids; the only stationary states for an electron are those quantized states that are not taken up by other electrons, and no states other than these can be occupied. We thus see that a solid or liquid made up of N identical particles has, if N is large, $2N$ possible states lying close together in bands that may be separated by gaps in which there are no quantized states, and which are, therefore, forbidden to the electrons. The band widths do not depend on the size of the crystal; the number of levels in a band increases in proportion to V, the volume of the crystal, but the separations between levels decrease appropriately.

The discrete atomic levels correspond, in a solid, to discrete bands with forbidden gaps between (Fig. 33). Closed groups of electrons in an atom or molecule usually become full bands in a solid or liquid; the substance acts as an insulator at the absolute zero. Examples are the atomic lattices of He, Ne, and argon, and molecular or ionic lattices. The group II elements (Be, Mg, Zn, Cd) have closed groups with two s electrons, though, and so should be insulators if the above were the only considerations; they are, in fact, metals and have unfilled bands. The reason is the overlapping of bands from different states; closed

electron groups in the atoms occupy only part of common bands in the solid.

The atomic levels do not merely split but, also, continuously change in energy, as atoms come together. The very band sequence may alter; a band derived from a high-energy atomic level may become lower in energy than a band derived from a lower-lying atomic level. The bands in a solid originate from atomic levels, but their energies, and sometimes their degrees of filling, may differ from those of the atomic levels.

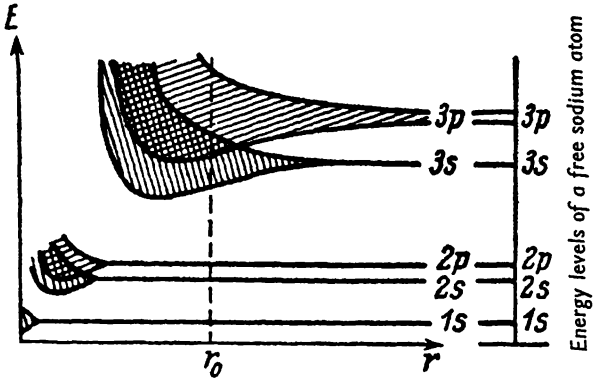


Fig. 33. Level splittings as atoms approach one another, and formation of the bands in a solid. The spacings between the levels are schematic only. r_0 is the atomic spacing in sodium crystals.

States of equal energy are to be found in atoms and molecules; such states are termed degenerate. The interactions between atoms in a solid may cause these degenerate states to have differing energies in the band system. The three p states in an isolated atom may give three distinct bands as the atoms are brought together. Degenerate atomic states may thus give non-degenerate bands in a solid.

Conduction electrons

The forbidden band divides the full ground-state band from the excited-state levels, which are free; any free state can be occupied by an electron coming from the full band or from an impurity atom. However, the free state cannot be reached by small steps because the intermediate energies would lie in the forbidden band. The energy source must supply an energy of at least ΔE_0 (the forbidden band width) at one step. Photons with $h\nu > \Delta E_0$ form a suitable source; so do the random thermal motions of the solid, although the mean energy of thermal vibration is much less than ΔE_0; kT is about 0·03 eV at room temperature, while ΔE_0 is usually 0·5–1·5 eV in a semiconductor. The

probability W_0 that an energy ΔE_0 will be passed on to a bound electron, for a mean energy kT, decreases as ΔE_0 increases in accordance with Boltzmann's law

$$W_0 \propto N_0 \exp\left(-\frac{\Delta E_0}{kT}\right), \tag{60}$$

where N_0 is the total number of electrons exchanging thermal energy with the lattice. In our case the electrons have different initial energies within the range covered by the full band; they need energies which exceed ΔE_0 by different amounts in order to reach the free or conduction band. ΔE_0 is only the difference between the least energy in the conduction band and the greatest in the full or valence band; the relation (60) thus cannot be used for all the electrons in the valence band.

Impurity atoms, or atoms of a compound in excess of the stoichiometric proportion may be able to provide electrons for the conduction band. The electrons from these atoms, which do not take part in forming the valence band, are found on or near their atoms, unlike the others, which are spread throughout the substance. These localized electrons may have energies other than those of the main lattice; their energies may lie in the forbidden band and near to the edge of the conduction band. The energy ΔE required to reach the conduction band is then less than ΔE_0.

If these impurity atoms have N electrons in all (and by impurity, in future we mean all atoms that are not part of the main lattice), the probability W that one will be freed is

$$W \propto N \exp\left(-\frac{\Delta E}{kT}\right).$$

Now $\exp(-\Delta E/kT)$ is the smaller, the larger $\Delta E/kT$. The larger ΔE_0 relative to ΔE at any given temperature, the less is W_0 relative to W. W_0 is much less than W if $(\Delta E_0 - \Delta E)$ is much greater than kT; the impurity levels are then virtually the sole source of electrons.

The electrons in the conduction band, no matter what their source, exchange thermal energy with the lattice vibrations and reside in the lowest energy levels. Every such electron has an equal chance of returning to a free state in the valence band or in the impurity levels. The number of such states n equals the number of electrons that have passed into the conduction band. The probability z that an electron will leave the conduction band is thus proportional to n^2. The system comes to an equilibrium when W equals z, while the free electron concentration n_0 produced by excitation from the valence band is given by the formula stated before

$$n_0 = 2\left(\frac{2\pi mkT}{h^2}\right)^{\frac{3}{2}} \exp\left(-\frac{\Delta E_0}{2kT}\right). \tag{61}$$

The ratio of n (electrons from impurities) to n_0 depends on $(\Delta E_0 - \Delta E)$ and on T. n/n_0 is very large when $(\Delta E_0 - \Delta E) \gg kT$; then the impurities alone determine the free electron concentration. At high temperatures, even if $(\Delta E_0 - \Delta E)$ is greater than kT, N is so much less than N_0 that n is negligible compared to n_0; the main lattice can then be taken as the sole electron source. If n/n_0 is not so large that we can neglect n_0, and not so small that we can neglect n, the free electron concentration is not given simply by $(n+n_0)$, since any free electron can go back to any vacant level. However, the two types of transition have different probabilities, and the formula for the free electron concentration must take account of all these factors.

Our simple analogies between the means by which electronic semiconductors and electrolytes dissociate must be refined. The impurity ions in electrolytes often differ in nature from the host ions, while the electrons in a semiconductor are all identical. The dissociation of the host and impurity ions can often be considered independently, while the electron concentrations are additive. Electrons derived from impurities or from the valence band can return to any vacant impurity or valence band level in electronic semiconductors. This somewhat distorts the intermediate range of the plot of conductivity against temperature (Fig. 26).

The electrons in a semiconductor have quantum numbers relating to parameters other than the energy. An electron in a given quantized state can enter only those free states allowed by the selection rules. Electrons in s states in the valence or impurity bands, for instance, can pass only to p states in the conduction band. This restriction does not apply to purely thermal transitions, since phonons can then assist; phonons have high momenta and can alter the quantum numbers of the electrons. There are no selection rules for ions in electrolytes.

Ions and electrons also differ in the conditions under which they can transfer energy to the lattice. The energy freed when an ion returns to an empty site at first raises the local temperature and then spreads throughout the crystal. The ion transfers energy to its neighbours via its damped oscillations. An electron which goes from a state of energy E_1 in a higher band to one of energy E_2 in a lower band or in local impurity levels has to give all its excess energy $(E_1 - E_2)$ to the lattice at once, because electrons cannot have energies lying within the forbidden band. The lattice can take up energy only as single quanta of energy $h\nu$, with the frequency ν close to the natural lattice frequency, i.e. in amounts of energy of about 0·03 eV only. Now $(E_1 - E_2)$ is often about 1 eV in recombinations, so some tens of phonons have to be emitted simultaneously. Quantum theory shows that such events are very improbable.

The energy could also be emitted as a photon with $h\nu$ equal to $(E_1 - E_2)$. This actually happens in luminescence, but far from all transitions from high to low energy states give such radiation.

Adirovich[1] has shown how this difficulty of radiationless transitions can be dealt with for electrons falling from free or excited states to impurity levels. The energy is transformed to energy of elastic deformation of the surrounding lattice, and then to thermal energy; he has shown that this process can conform to the requirements of quantum theory.

The energy freed when an electron is trapped by an F-centre can be treated similarly. An F-centre is an electron replacing a negative ion in the lattice of an ionic crystal. Pekar[2] has given the theory of the effect, and has calculated the electron recombination probability for cases where the electron wanders close enough to a trap, e.g. to a colour centre in an alkali halide;[3] the agreement with experiment is good. The capture probability is large compared with the collision probability. However, not all encounters end in trapping; many collisions are needed for an electron to be attached to a chlorine or oxygen molecule, for example; only one collision in hundreds of thousands is successful. Similar effects must occur in a solid; the combination probability is the product of the collision and capture probabilities. Here instantaneous proximity and excess energy are not enough to cause an electron to pass from a free to a quantized level. Electrons and holes seldom combine directly; one or other is first trapped on an impurity or a lattice defect.

Sosnovskii[4] has shown that a recombining free electron can pass its excess energy to another free electron that then loses its excess kinetic energy gradually to phonons as it descends towards the bottom of the conduction band.

The Adirovich–Pekar theory explains free carrier combination at fixed centres, but there is no adequate theory explaining how non-localized electrons and holes of similar mobilities recombine. We can make the likely assumption that all the free electrons have equal recombination probabilities which are proportional to the numbers of free levels in the valence band and among the impurity levels. Any future theory must give the coefficients of proportionality for particular cases.

Surface recombination is as important as is bulk recombination; the recombination rate at the surface may be high if the state of the surface is suitable. Surface treatments (polishing, etching) can change the recombination probabilities near the surface greatly. Recombination at junctions between semiconductors is also important, but no exact theories of this effect exist as yet.

The quantum rules cause electron transitions to the conduction band to have some special features not found with ionic dissociations. Atoms

and ions obey statistics different from that for electrons, namely Bose–Einstein, which allows of any number of particles in one cell in phase space; electron statistics allow only two electrons with opposed spins per cell.

We cannot understand correctly how electrons behave in solids without allowing for the quantum laws governing their motion, and for the effects of the local fields produced by the atoms or ions. Only the mean interactions defined by the polarizability and dielectric constant were taken into account when we considered the condensation of a rarefied gas to a liquid or solid. We did not point out that electrons move freely in a semiconductor because they have a wave nature. Electrons of wavelengths given by Bragg's law, $\lambda = 2a \sin \theta$ (where a is the distance between adjacent atoms and θ is the angle between the direction of motion and the lattice planes), are totally internally reflected and cannot, therefore, continue to move in the same direction at the same speed; the travelling waves become standing ones. This is why electrons are forbidden certain energies and speeds in a semiconductor; in accordance with the discrete energy spectrum, forbidden bands occur.

The two arguments give similar pictures of the energy spectrum of the electrons in a semiconductor. The actual spectrum, of course, cannot depend on the method of deduction, since we are dealing with the same electrons and the same medium.

The interactions of an electron with the lattice, and the fields that result, must be taken into account when we consider the level splitting that occurs as the atoms are brought together; it is not enough to consider the quantized state of the electron in the isolated atom. The two theories of the energy spectrum do not contradict but supplement one another where they are capable of giving exact results.

We turn now to the properties of electrons in semiconductors. The theory is much simplified by replacing the semiconductor by an ideal crystal lattice. This gives the only highly developed theory of semiconductors. However, semiconductors are not necessarily crystalline, and are not obliged to have long-range order. Electronic conduction has been demonstrated in amorphous bodies lacking regular crystalline structures and having only short-range order. Examples are liquid metals, fused semiconductors, sodium and potassium dissolved in liquid ammonia, amorphous selenium and semiconducting glasses.

While the conditions of electronic motion are different for amorphous and for crystalline materials, the discrete energy spectra (especially the forbidden bands), thermal transitions between levels, recombination, diffusion and statistical equilibria are the same for both. The interatomic bonds, the atomic arrays, and the consequent quantized states, determine the properties both of amorphous and of crystalline semiconductors.

There is as yet no proper theory of amorphous semiconductors, but only a few first steps towards one. There can be no doubt, however, that, as amorphous semiconductors come into technical use, such a theory will be produced. Some vitreous semiconductors are in fact already in use.

In the next sections we shall first deal with the motion of free electrons *in vacuo*, and then with the changes caused by the field periodicity in a crystal. Special attention will be given to effective masses and to hole and electron conduction mechanisms.

The laws of motion for free electrons

The momentum P is used to describe the motion of an electron, considered as a particle, in free space; the energy E is also used for this purpose. E and P are related by

$$E = \frac{P^2}{2m}.$$

P is given in terms of the wavelength λ by wave mechanics

$$mv = P = \frac{h}{\lambda} = \frac{2\pi\hbar}{\lambda}.$$

Here $\hbar$ is $$\frac{h}{2\pi} = \frac{6{\cdot}62 \times 10^{-27}}{2\pi} = 1{\cdot}06 \times 10^{-27} \text{ erg sec},$$

while $$\frac{2\pi}{\lambda} = k \tag{62}$$

is the wave number, which is related to P by

$$P = \hbar k \tag{63}$$

and which, like P, defines the direction of motion. E, the kinetic energy, can be put in terms of k and of the mass m

$$E = \frac{\hbar^2 k^2}{2m}, \tag{64}$$

whence $$\frac{dE}{dk} = \frac{\hbar^2}{m} k \tag{65}$$

and $$k = \frac{m}{\hbar^2} \frac{dE}{dk}. \tag{66}$$

The speed v is also defined by k and m

$$v = \frac{P}{m} = \frac{\hbar k}{m}$$

or, from equation (66),

$$v = \frac{1}{\hbar}\frac{dE}{dk}. \tag{67}$$

The principle of indeterminacy shows that an electron of strictly defined P or k is not localized in space.

The conditions of motion of an electron in a semiconductor

A free electron moving in a field-free region has a constant velocity v; an electron in a crystal keeps the same speed only on the average, or over short distances. Classically speaking, the speed falls and rises again as the electron passes through the potential barriers between the atoms or ions. Real electrons have a wave nature, and their instantaneous velocities are indefinite by an amount $\Delta v = h/m\Delta a$, where Δa is of the order of the lattice constant. Putting $h = 6 \times 10^{-27}$ erg sec, $m = 9 \times 10^{-28}$ gm and $\Delta a = 3 \times 10^{-8}$ cm, we have Δv equal to 10^8 cm/sec. Δv is thus larger than v itself, which is about 10^7 cm/sec for free electrons. The instantaneous velocity v has no meaning for a semiconductor; the mean speed over distances many times the atomic spacing is alone significant.

All the electrons in a semiconductor should be considered as forming a single system distributed over the energy levels. The electrons are then not localized and have finite probabilities of being found anywhere in the body. An electron can move without being scattered if its energy and momentum lie in the allowed ranges, and if the field is strictly periodic.

However, in practice the electrons in the deep-lying K, L, etc., shells differ greatly from the valence electrons. The first are bound very strongly to their nuclei, and their potential barriers are thousands of eV high; their wave functions are limited to small volumes round the nuclei, and those of the K and L electrons on different nuclei do not overlap. Their electrostatic interactions are negligible relative to the forces binding them to their nuclei. Their energy levels are hardly split and form spectral bands almost as narrow as the lines in the isolated atoms. The K and L electrons are localized near their nuclei.

The probability that such an electron will be found in any part of the crystal is finite, but the times needed are enormously long. This means that such electrons, in fact, remain near their nuclei and take no part in conducting a current or in diffusion, even if their shells are not full. If their shells are full, their bands in the solid are also full; this is a second reason why such electrons cannot take part in diffusion or in current flow. The situation is different for the valence electrons, which have wave functions that overlap; they are not localized.

The free electron and hole concentrations are often so small (10^{15} to 10^{18} cm^{-3}) that we can assume the individual wave functions to be independent and hence the positions of electrons also.

We now examine the movement of free electrons in a solid. The wave function ψ of an electron moving *in vacuo* along the x axis is

$$\psi(x) = A \exp(ikx) \tag{68}$$

where A is a constant not dependent on x. The wave function for motion in a periodic field can be put in a like form by making A a periodic function of x, the period being that of the lattice. A also depends on the wave vector k if the body is anisotropic. ψ can thus be put in the same form as for a free electron

$$\psi_k(x) = A_k(x) \exp(ikx) \tag{69}$$

and for a given k

$$A_k(x) = A_k(x + na) \tag{70}$$

where a is the lattice constant and n is an integer. The wave functions are harmonic waves, which are amplitude-modulated with a period equal to the lattice constant.

Now the atoms or ions in the lattice lie at its nodes, and the K, L and M electrons are bound to their nuclei; the valence electrons, except for the one we are considering, fill all the rest of the space. Equation (69) then satisfies the law of quantum mechanics that describes the motion in a periodic field. The assumptions made here are quite natural for a semiconductor; the heavy atoms or ions change their positions only slowly, and so can be taken as fixed when we deal with the fast-moving electrons. The enormous number of randomly moving electrons may conveniently be replaced by a charge spread over the field produced by the atoms in their instantaneous positions; the effects from our one electron can be neglected. This electron can alter the charge distribution to distances where its field becomes comparable with that of the lattice. Now the mobile electrons in semiconductors are usually comparatively few (some 10^{14} to 10^{18} per cm^3), so they are tens or hundreds of atomic diameters apart on the average. We can thus assume that the motion of one such electron will not appreciably alter the distribution of the others.

Although it is not strictly correct to consider one electron moving in the field caused by the other electrons and by the ions, equation (69) is a good approximation for an electron in a crystal and satisfies the wave equation

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi + V_p(\mathbf{r}) \psi \tag{71}$$

where ∇^2 is the sum of the derivatives with respect to the three coordinates and $V_p(\mathbf{r})$ is the potential energy. This one-electron theory

gives nearly the same results as does the more rigorous but vastly more complex many-electron one.

The probability of finding the electron at any point in space equals the product of ψ by its complex conjugate, i.e. the square of the modulus of A. Equation (68) shows that $|\psi_k(x)|^2$, for an electron *in vacuo*, is independent of x if A is constant. The electron is equally likely to be at any point in space; the motion is described by a plane wave of period defined by k.

The laws of motion of electrons in a periodic field

A is a periodic function of x when the electron moves in a crystalline lattice, so the probability of finding the electron at any point is also a periodic function of x. Positions differing in x by multiples of a are equally probable; we cannot say at which of them the electron will be at any given time. Only the positions within a period are physically distinguishable; we cannot say in which of the many unit cells a given electron may be.

The problem of the motion of an electron in a solid body has not yet been solved in a general form, but a model in which the field produced by (1) the instantaneous ion or atom distribution, which may, subject to corrections for thermal motions, be taken as periodic, and (2) the electron distribution implied by (1), gives an adequate approximation. The electron as it moves is acted on by the periodically varying internal field and by the external electromagnetic field. If we average the parameters describing the motion over several periods, we can describe the motion due to the external field by equations of the same form as those for an electron *in vacuo*, although the interactions with the body are much more complex.

Equation (67) relates the instantaneous values of the energy, wave vector and speed but does not apply to a solid, unless by v we understand the speed averaged over a free path, and not the instantaneous value; this mean speed may, by analogy with equation (67), be derived for steady-state motion and a free electron from

$$\bar{v} = \frac{1}{\hbar}\frac{dE}{dk}. \tag{72}$$

Here again the energy E can be given in terms of P^2; we term P the quasi-momentum

$$E = \frac{P^2}{2m}$$

and $P = \hbar k$, so v can be put as a function of k,

$$\bar{v} = \frac{\hbar k}{m}.$$

The acceleration is

$$\frac{d\bar{v}}{dt} = \frac{1}{\hbar}\frac{d}{dt}\left(\frac{\partial E}{\partial k}\right) = \frac{1}{\hbar}\frac{\partial}{\partial k}\left(\frac{dE}{dt}\right). \tag{72a}$$

The change in E in a time Δt is ΔE and is the work done by the force F. The work done by F in a distance Δx equal to $\bar{v}\Delta t$ is $F\bar{v}\Delta t$

$$F\bar{v}\Delta t = \frac{dE}{dt}\Delta t.$$

Here Δx is a length which covers several lattice periods in the x direction. Substituting for dE/dt in equation (72a) and replacing $\partial E/\partial k$ from equation (72), we get

$$\frac{d\bar{v}}{dt} = F\frac{1}{\hbar^2}\frac{\partial^2 E}{\partial k^2}. \tag{73}$$

The effective mass

The equation of classical mechanics relating dv/dt to F is

$$\frac{dv}{dt} = \frac{F}{m}. \tag{74}$$

Comparing this with equation (73), we see that $(1/\hbar^2)\,\partial^2 E/\partial k^2$ is analogous to $1/m$ (the reciprocal of the mass m). There is no reason for saying that this mass must equal the mass of a free electron, because $(1/\hbar^2)\,\partial^2 E/\partial k^2$ is constant only if E is a quadratic function of k. So, if we put

$$\frac{1}{\hbar^2}\frac{\partial^2 E}{\partial k^2} = \frac{1}{m^*} \tag{75}$$

and call m^* the effective mass, we assign to this term none of the properties of a mass other than that it is the coefficient of proportionality between the external force and the acceleration which satisfies

$$E = \frac{\hbar^2 k^2}{2m^*}.$$

We shall see how this applies to an electron near the lower edge of the conduction band, or to a hole near the top edge of the valence band. m^*, unlike a proper mass, does not define the energy, or the weight, or the inertia; m^* is a function of direction in anisotropic bodies and is then expressed by a tensor, whereas ordinary mass is a scalar. $d\bar{v}/dt$ and F may not even be in the same direction.

We are justified in using the effective mass by the convenience of the laws of classical mechanics; provided that we remember that the effective mass is not a mass in the usual sense of the word, but only an abbreviation for $\hbar^2/(\partial^2 E/\partial k^2)$, the term is quite applicable.

We must not forget that in our descriptions of electrons in solids we use quantities which only resemble those for a free charge. We should consider it fortunate that our use of a mean speed has enabled us to refer the motion of an electron in a periodic field to the free motion of the same charge e, just replacing the true mass m_0 by the effective mass m^*. The electron then satisfies equation (71), the basic equation of quantum mechanics; the wave function is equation (69). The electron retains the same mean velocity in the lattice if the periodicity of the latter is not disturbed; the mean acceleration caused by an external field $\mathscr{E}$ is $d\bar{v}/dt = \mathscr{E}e/m^*$.

This device is of great help in giving us a clear picture of the processes in semiconductors, but we must not forget our assumptions; we have simplified the electric field within the solid, neglected electron interactions, assumed a periodic lattice, and considered thermal agitation as a weak perturbation.

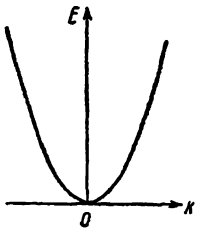


Fig. 34. Relation of kinetic energy E to wave vector k for completely free electrons.

Let us plot the kinetic energy E as a function of k for a free electron in one dimension. Equation (64) gives

$$E = \frac{\hbar^2}{2m} k^2.$$

E is essentially positive and is set as ordinate, while k, which can be positive or negative, is set as abscissa. Equation (64) then gives a parabola (Fig. 34) with the curvature at the origin given by

$$\frac{\partial^2 E}{\partial k^2} = \frac{\hbar^2}{m} \tag{76}$$

which depends on the mass; for a free electron m is 9×10^{-28} gm. m is constant at all values of E and k (if $E/c^2 \ll m$, where c is the speed of light); all possible values of k and positive values of E are open to a free electron. An electron in a field of constant potential (Sommerfeld's theory of metals) gives a curve of the same shape.

We now consider electrons in a solid. If the perturbation caused by the periodic field is slight, we say the electrons are weakly bound

to the lattice; otherwise we say they are strongly bound. The binding is strongest when the electron is localized on a particular atom. We shall in future deal only with weak binding and non-localized free electrons. Not only does the curvature of E as a function of k, which defines m^*, have a different value from that for a free electron *in vacuo*; the energies open to the electron fall into discrete bands with forbidden bands between. The electrons at the edges of the allowed bands must have zero velocity. But v is equal to $(\partial E/\partial k)/\hbar$, so $\partial E/\partial k$ is zero at the band edges; hence the E against k curve becomes horizontal at these points. Fig. 35a shows E plotted against k for an s band, and Fig. 35b the same plot for a p band. Every such band derived from an atomic state contains $2N$ states, where N is the number of atoms; every one of these $2N$ states has its own E and k. The largest k in any band is defined in terms of a, the lattice constant along the direction of k.

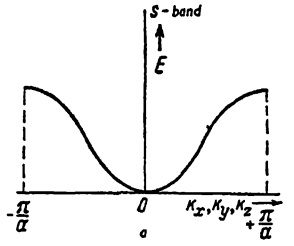


Fig. 35a. Relation between energy and wave number for an s band.

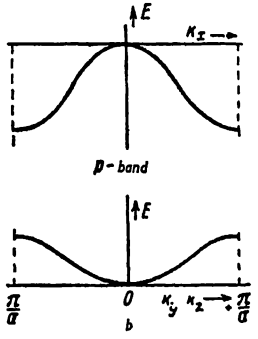


Fig. 35b. Relation between energy and wave number for a p band.

The wave nature of electrons appears in interference effects; the waves weaken rapidly, and then are completely reflected, as $2\pi/\lambda$ equal to k approaches the value required to satisfy Bragg's law

$$\lambda = 2a\cos\theta.$$

Hence the value of k which gives complete reflection is

$$k = \pm\frac{n\pi}{a} \tag{77}$$

where n is an integer.

The wave is totally reflected whenever k is $\pm n\pi/a$; electrons with these values of k cannot move in the crystal; the travelling waves become standing ones. Equation (77) sets limits to the possible values of k. Within a band k varies from $-\pi/a$ to $+\pi/a$; n does not affect the value of k or the momentum.

The speed, and hence k, change continuously if the force acting on the electrons (e.g. from an electric field) is constant in direction. There is, however, a limit to k; when a value of $n\pi/a$ (i.e. a band edge) is reached, k jumps back to its value at the other edge of the band (a change of $2\pi/a$), while P changes by $2\pi\hbar/a$. The electron is as it were reflected from the band edge.

The band edges are dependent on the direction of k; the limits of k vary with direction and correspond to the way the lattice spacing a varies. The more strongly the valence electrons are bound, the higher the barriers between the atoms, the narrower the allowed bands and the wider the gaps between them. The widths of these gaps also depend on direction, but there will generally be some range ΔE_0 in which no motion is possible; this is the forbidden band. Outside this band there may be energy levels which allow of motion in some directions but not in others.

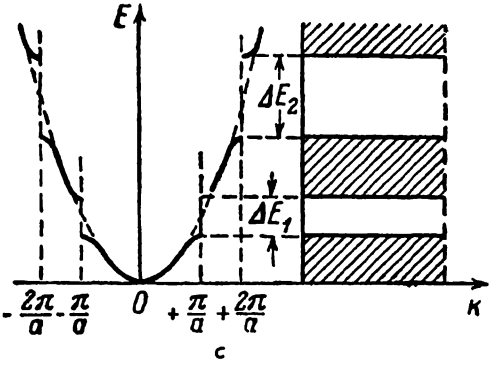


Fig. 35c. Typical $E(k)$ curve for electrons which are not entirely free. The broken line (parabola) is the $E(k)$ plot for completely free electrons.

The gaps in the allowed energies in certain directions may vanish, if the gaps themselves are small; there will then be no forbidden band, although in any given direction k_j there will be energy discontinuities at $\pm\pi/a_j$.

The gaps in one set of split levels may be covered by the split levels of another set, if the two sets originate from different states of the free atom; e.g. in copper the set deriving from the $3d$ states overlaps the $4s$ system. The elements in group II in the periodic table are metals only because two sets of levels overlap in this way.

Crystals formed by normal molecules derive their bands from the splitting that occurs in the molecular levels as the molecules are brought together; the bands are either entirely full or entirely empty.

The value of the integer n specifies the band; the first extends from $k = 0$ to $k = \pm\pi/a$, the second from $\pm\pi/a$ to $\pm 2\pi/a$, etc. The properties of the electrons repeat at intervals of $2\pi/a$, however, so we need consider the first period only. While k increases continuously (or repeats), E changes by steps of ΔE_0 between bands, just as when an electron jumps between atomic levels. These general arguments enable us to represent E as a function of k for any given direction in which the lattice constant is a (Fig. 35c). Fig. 35c is plotted in the same way as Fig. 34. The bands produced by splitting the atomic levels are shaded; the forbidden bands are unshaded.

So far we have dealt with motion having a definite direction for k only. Each band contains states with k ranging from $\pm n\pi/a$ to $\pm(n-1)\pi/a$, and the energy gaps between the bands are ΔE_n. The values of a, and hence the limiting values of k and E, depend on direction in an anisotropic body.

The Brillouin zones

No drawing analogous to Fig. 35c can be made for three dimensions, but we can project k in the Cartesian system and can represent the surfaces of equal energy in these coordinates. If k is small (far from π/a), the anisotropy in a crystal causes the surfaces of equal energy to be ellipsoids, but the equipotential surfaces become distorted as π/a is approached, for some directions, and even if k is far from π/a, for others. We must therefore first decide on the boundaries of each band of levels in the three-dimensional k_x, k_y, k_z space for the lattice. These boundaries are polyhedra and are called the Brillouin zones.

In the simplest lattice, where the atoms lie at the corners of cubes, the boundaries of the first zone are given by $k_x = k_y = k_z = \pi/a$, so the first zone is a cube of side $2\pi/a_0$, where a_0 is the edge of the elementary cube. The second is bounded by a rhombic dodecahedron (which encloses the cube) with its opposite faces $2\sqrt{(2)}\,\pi/a_0$ apart (Fig. 36a). All states in the first zone have their k_x, k_y and k_z within the cube, which has its faces distant π/a_0 from the origin. The states in the second zone fill the space between the cube and the dodecahedron (all zones of cubic crystals are of equal volume in the k_x, k_y, k_z system). The energy changes stepwise by ΔE_0 on passing through a boundary; ΔE_0 itself depends on the direction of k. Fig. 36b illustrates the first and second Brillouin zones for a body-centred cubic lattice, and Fig. 36c does the same for the face-centred cubic case.

Holes and electrons

Fig. 37 shows two-dimensional equipotential curves in the k_x, k_y plane within the square bounding the first zone, and outside the square

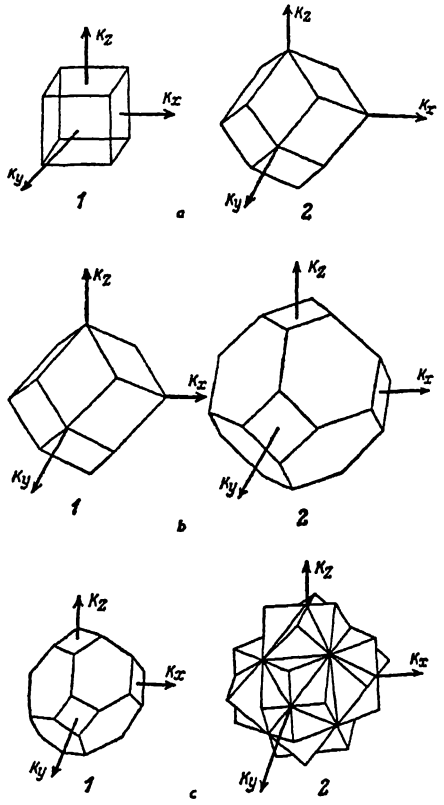


Fig. 36. Shapes of the first (1) and second (2) Brillouin zones: (a) simple cubic lattice, (b) body-centred cubic lattice, (c) face-centred cubic lattice.

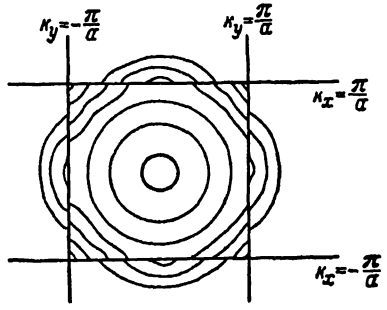


Fig. 37. Curves of equal energy in the k_x, k_y plane. The discontinuities occur only at the lines corresponding to zone boundaries.

(in the second zone). Fig. 35 shows how arbitrary the effective mass concept is. Only at the lower edge of each band does the E against k curve resemble a parabola, which corresponds to a free electron of constant mass. Higher up we reach a point with $\partial^2 E/\partial k^2 = 0$; the curvature then changes sign, and $\partial^2 E/\partial k^2$ becomes negative. At the top edge we again get a parabola, but this time inverted, with opposed curvature. Thus in every band the effective mass given by equation (74) starts at some positive value $+m^*$ at the bottom, rises to infinity, changes sign, and tends to $-m^*$ at the top edge. All resemblance to the motion of a free electron *in vacuo*, where the mass cannot be negative, is lost. The physical meaning of this is that there is only a limited region where it is possible to dispose of the periodic field by choosing an appropriate effective mass.

We can describe the motion of an electron in a crystal in terms of the laws for a free electron because E can be represented as a function of k, as in a vacuum, by

$$E = \frac{\hbar^2 k^2}{2m^*}. \tag{64a}$$

m^* cannot be derived from equations (66) and (75), if we still wish to call it a mass, if the $E(k)$ relation is other than that of equation (64a). Equation (64a) is satisfied near the top edge of the valence band, so we can speak of some effective mass m^* given by

$$\frac{1}{m^*} = \frac{1}{\hbar^2}\frac{\partial^2 E}{\partial k^2}$$

but this mass is negative. We now examine how an electron behaves in a state which gives it a negative effective mass. An electric field in the same direction as the velocity retards the electron, and does not accelerate it. The speed, defined as $(1/\hbar)\partial E/\partial k$, decreases as E increases.

We now examine how such states occur in semiconductors. If a band contains far more states than electrons, the latter will lie in the low-energy states at the bottom of the band, where m^* is positive. If all the states are filled no external force can cause any effect, since the electrons must retain their states. However, if just a few of the states are empty, say because some electrons have gone to a free band, or because the valence electrons are rather fewer than the number of states, only the states of highest energy can be free. We thus have empty states near the top edge of the band. The properties of the electrons in such a band can be specified in terms of the behaviour of these empty states.

Now electrons tend to enter vacant states of lower energies, so these empty states ("holes") receive electrons from higher energy states which leave holes behind them. The effect is as if a hole passes from a low

state to a higher one; if, moreover, the change in the energy of the electron assembly is ascribed to a corresponding change in the energy of the hole, we must say that holes are of higher energy in the deeper-lying electron levels. The energy of a hole increases from the top of a band downwards.

Equation (73) gives the acceleration of a free electron by an electric field; electrons near the top of a band have $\partial^2 E/\partial k^2 < 0$, and thus $\partial \bar{v}/\partial t < 0$ if $F > 0$. The force thus retards such electrons; it is difficult to use the ideas of classical mechanics here. However, we consider the effect as one in which the electron moves into a state of lower velocity and thereby frees a state which corresponds to a higher velocity. We then take the freed state as a hole and assign to it a positive charge, finding that the field drives the hole in the appropriate direction and increases its energy. There is nothing unusual in such a result.

The vacant states in a band are filled by electrons which have vacated their original states. The vacant states (holes) move within a nearly full band in just the same way as positive charges would if they had positive effective masses

$$m^* = \hbar^2 \left| 1 \Big/ \frac{\partial^2 E}{\partial k^2} \right|. \tag{78}$$

The motion of an electron in a solid is described in terms of the effective mass, which is similar to, but not the same as, the mass of a free electron. The hole concept likewise enables us to refer the properties of a band with but few vacancies to classical laws; a hole is like a positive charge. In both cases the complex laws for real electrons are replaced by simple analogies; a clear picture is created, and the laws of motion for free electrons can be transferred to electrons in a semiconductor. We must not forget that our negative electrons and positive holes, which obey the ordinary laws of electrodynamics, although with differing effective masses, are approximations which hold good only in limited regions of quantized states, i.e. near the bottom edge of the conduction band and near the top edge of the valence band. This corresponds in Fig. 35 to the small region of the E against k curve, near the edge of the forbidden band, which can be represented as parabolic. In the ranges of E and k for which the $E(k)$ plot differs but little from a parabola, we can take the distance from the band edge, which is a measure of the energy, as a quadratic function of k

$$E = \frac{\hbar^2}{2m^*} k^2,$$

and

$$\frac{1}{m^*} = \frac{1}{\hbar^2} \frac{\partial^2 E}{\partial k^2}$$

is constant and analogous to a reciprocal mass. But what does the effective mass mean in the region of $E(k)$ where $\partial^2 E/\partial k^2 = 0$, yielding $m^* = \pm\infty$?

If we were to try to apply these concepts to a whole band we would have to say that the mass of an electron tends to infinity as we pass upwards towards the inflection from the bottom of the conduction band, and that the mass of a hole does the same as we pass downwards from the top of the valence band. No finite force would appear able to cause the electron or hole to pass through such a boundary. But this is not so; electrons can be removed from any level in a band. The effective mass, defined as the ratio of the force to the acceleration, is meaningful near the appropriate band edges only, since only here can the kinetic energy be represented as the square of the momentum. $\hbar^2/(\partial^2 E/\partial k^2)$ ceases to resemble a mass if $E(k)$ takes any other form.

The arbitrary natures of the effective masses and of the assignment of a positive charge to a hole can be demonstrated by measurements of the gyromagnetic effect (i.e. measurements of the field caused by rotating a metal rod rapidly about its axis), or by measuring the current pulse produced by suddenly stopping a rotating current-carrying coil. The magnitudes and signs of the fields or voltages give us the ratios of charge to mass and the signs of the charges. Barnett's accurate measurements show that: (1) the charges are always negative, even for metals for which the Hall effect and the thermoelectric power indicate hole conduction; (2) the charge carriers differ little in mass between metals and are always of masses close to the mass of a free electron, even though the effective masses of the carriers given by the thermoelectric, magnetic, thermal, optical or electrical properties differ greatly from those of free electrons.

This shows that the effective mass has no relation to the inertial features inseparable from mass proper; it also does not have another basic feature of mass proper, namely weight. A free electron of effective mass m_-^* and a hole of effective mass m_+^* are formed when an electron enters the conduction band from the valence band; the weight of the body remains the same, and does not increase by m_+^*. Although many of the electrons, particularly the innermost ones, are bound in very narrow bands and have high effective masses, the weight of the body is unaltered. The sum of the effective masses would often have been high enough to make up a large and easily measurable fraction of the total weight.

Hence m^* equal to $\hbar^2/(\partial^2 E/\partial k^2)$ is like an ordinary mass only in relation to the movements of electrons and holes caused by external fields if the particles lie near the appropriate edges of their bands. We may well wonder whether the concepts of effective mass and a hole

are useful if their ranges of application are so narrow. Their use is justified by the fact that in all cases so far examined we have just the conditions required (regions near band edges). In practice, all effects of importance in semiconductors can be described very accurately in terms of effective masses and positive holes. Difficulties, which have so far not been overcome, would be encountered if the approximation were rejected and all the electrons had to be taken into account in every particular case, since quantum mechanics has not yet solved the problems of the dynamics of many-body systems. We therefore take the electrons at the bottom of a band, and the few vacant levels near the top of a band, as having negative and positive charges respectively, and as having effective masses m^* defined by

$$\frac{1}{m^*} = \frac{1}{\hbar^2}\frac{\partial^2 E}{\partial k^2}.$$

Values of m^* given by this formula for the central areas of bands cannot be thought of as being like masses in classical mechanics.

Holes and electrons differ in more than their charges. They lie in different bands, the holes in the valence band, the electrons in the conduction band. The bands differ in width, and the narrower the band the larger the effective masses; the conditions of motion therefore differ, and temperature and pressure affect holes and electrons differently.

Atomic levels which lie close together may be so split in the solid as to give bands that overlap; there is, for example, a second narrow band overlapping the normal band in tellurium. Both bands can give holes, but the corresponding masses differ. Traces of holes of different (larger) masses are also found in other cases.

The effective mass is determined by the manner in which the charge carrier interacts with the lattice fields, which have the lattice symmetry, but which may be very complex. The effective mass, unlike the inertial or gravitational mass, is not a scalar but a tensor, i.e. is dependent on the crystallographic direction. It has one value for acceleration along the direction of motion (longitudinal mass) and another for acceleration normal to the line of motion (transverse mass).

Carrier mobilities

The holes and electrons in a semiconductor move as would positive and negative charges equal in magnitude to the free electron charge with certain definite (but different) effective masses. The Hall constant is thus negative if we have electrons near the bottom of the conduction band, or positive if the holes near the top of the valence band carry the current. The thermoelectric power shows a similar change of sign. The laws of motion for the electrons and holes are like those for free

electrons if the effective masses are used, and the same equations apply. One important restriction must be observed in using these laws with a semiconductor, however. We can speak of free motion only in the ideal non-existent case of an unbounded ideally perfect crystalline semiconductor at the absolute zero. In real materials the obstacles to motion are the grain boundaries (in polycrystalline materials), the foreign or excess atoms or ions, thermal fluctuations, and the electric fields produced by such fluctuations. The equations that describe the relationship between the mean speed v and an external force F can be applied only to short lengths of the path l (the mean free path), or to the corresponding short times of relaxation τ.

Equation (73) shows how the speed varies during a time τ; the changes are small. The mean kinetic energy E is $3kT/2$ if the electron concentration is well below the limit of degeneracy, as is usual in semiconductors; its value at room temperature is about 0·03 eV. Even a field of 10^4 V/cm adds only 10^{-3} eV if l is 10^{-7} cm; this energy is $\frac{1}{30}$ of the thermal energy. The speed rises appreciably only if $l \gg 10^{-7}$ cm, or at fields of 10^5 V/cm or more.

The electrons thus move on an average at random with almost the same speeds whether or not an electric field is applied; the entire body of electrons merely drifts rather slowly at some mean rate which is proportional to the field. The ratio of this mean rate to the field strength is termed the mobility. Experience shows that Ohm's law applies to most semiconductors up to fields of 10^4 V/cm, so we can say that up to this limit the mobilities are field-independent. Our results show that even when Ohm's law does not apply, the mobilities of artificially produced carriers (e.g. photoelectrons) which differ in no way from free charges produced thermally, are often the same as before. The deviations from Ohm's law are caused by increases in the numbers of the free electrons.

Atomic lattices should give strong-field mobilities varying as $\mathscr{E}^{-\frac{1}{2}}$, where $\mathscr{E}$ is the field. Ionic crystals give mobilities which are field-independent up to several thousand volts/cm. Stronger fields alter the mean speeds appreciably, and we should expect that the mobility would change with the speed, and hence with the field.

Methods of measuring mobilities and effective masses

There are several methods of measuring the mobility u. The Hall constant gives n if the carriers are of one sign only; in atomic and molecular lattices, where the mean free path l is energy-independent,

$$R = \frac{3\pi}{8nec}. \tag{79}$$

At electron concentrations well above the degeneracy level

$$R = \frac{1}{nec}. \tag{13}$$

The factors which affect the scattering also alter R, but R is always given by A/nec, where A takes a value dependent on the scattering mechanism. We can calculate u if σ has been measured, since $\sigma = neu$.

$$R\sigma = \frac{3\pi}{8c}u, \tag{80}$$

or, in general,

$$R\sigma = Au/c.$$

We can also measure u by producing additional carriers photoelectrically or with a strong field, using a small area of the semiconductor and short times. The region of increased carrier concentration is moved at a rate v by an electric field $\mathscr{E}$; $v/\mathscr{E}$ gives us the mobility. This is the basis of the drift method of measuring mobilities.

The two methods give compatible results; u varies very widely, from 0·01 to 10^4 cm^2/V sec. The mobility is related to the effective mass; the greater m^*, the less u. In atomic lattices $u \propto (m^*)^{-\frac{5}{2}}$; in ionic lattices $u \propto (m^*)^{-\frac{3}{2}}$. u usually falls as T rises, or when impurities are added, though cases are known where u rises on the addition of impurities that remove lattice defects. u depends on the crystallographic direction in anisotropic crystals. The electrons and holes in a semiconductor may have differing mobilities, and differing effective masses. Ge has a ratio u_-/u_+ of 1·5 to 2, while InSb gives a ratio of 80.

The effective masses vary greatly, like the mobilities. The unit can conveniently be made the mass m_0 of a free electron *in vacuo* ($9{\cdot}1 \times 10^{-28}$ gm); m^*/m_0 (or m^* measured in these units) may range from well above unity (TiO_2, WO_3) to well below unity (Ge, InSb, HgSe). It is not quite so simple to measure m^* as it is to measure u; m^* can be found from the thermoelectric power combined with a knowledge of the electron concentration, for instance. The results are often divergent; the factors which affect m^* in any given crystal are as yet not clear. Resistance measurements coupled to measurements at high frequencies in magnetic fields (which are dealt with in the next chapter) give the most data on m^*.

Additions to the electron theory of metals

The above arguments show that we must reconsider the statistical electron distribution in a metal. We neglected the Brillouin zones and the discrete energy spectrum in our treatment of the quantum statistics of electrons in metals. The simple theory agreed well with experiment

because in metals we have an unfilled band with energies such as to lie far from the band edge; the surfaces of equal energy could therefore be taken as spherical in k_x, k_y, k_z coordinates without much error, the bands could be neglected, and we could take as energy $\hbar^2 k^2/2m$. We also made some other assumptions.

1. Electrons lying a distance about kT from the Fermi level are mainly responsible for all the effects of interest (thermal and electrical conduction, Hall effect, etc.). The electron mass in the formula may differ from that for a free electron, m_0, and should therefore be taken as the effective mass, m^*. In by no means all metals is m^* close to m_0.

2. In some metals (e.g. W, Mo, Zn) the bands are, we now understand, almost filled with electrons; holes carry the current.

3. The bands overlap in the alkaline-earth metals; it is very doubtful if the surfaces of equal energy are spherical near the Fermi level in such metals.

4. Hall effect measurements indicate that electrons and holes both carry the current in some metals (Pb, Sn, Pt). The energy levels here lie not far from the zone boundaries. Our theory does not apply to such metals.

There is no metal (except bismuth) for which there is a complete theory which allows exactly for the dependence of E on k_x, k_y and k_z (i.e. for the exact shapes of the surfaces). We have only some general ideas which explain qualitatively the effects found in metals; the deductions made agree with those from semiconductor theory. Though the latter originated from the theory of metals, it could not be bound by the assumptions that electrons in solids moved like free electrons, or that Fermi statistics were applicable. These assumptions were good enough for metals.

In semiconductors we have electrons and holes near the forbidden band, and electron transitions between levels that are far apart; all the quantum level systems must thus be studied in detail if semiconductors are to be understood. The variety of effects and the numerous different technical uses have forced us to develop a very much more detailed theory for semiconductors. Semiconductor theory has thus shown up some new aspects of electrons in solids and liquids.

Semiconductor theory now influences the theory of metals, especially of alloys, many of which on closer study are, in fact, found to be semiconductors.

The distribution of the electrons in a semiconductor among the quantum states

To find the distribution of the electrons in a semiconductor among the quantum states specified by the wave numbers k_x, k_y and k_z, we

have first to find the distribution of levels within a band, i.e. to find $n(E)$ as a function of k or of E, and then to find the probability w that a given level is filled. At equilibrium there will be $wn(E)\,dE$ electrons in the levels between E and $(E+dE)$.

The first problem is very difficult to solve in a general form. The simplified theory of metals can be used if the bands do not overlap, and levels near the bottom of a band are considered; here $n(E)$ is given as a function of E, as in equation (23), by

$$n(E)\,dE = 4\left(\frac{2m^*}{h^2}\right)^{\frac{3}{2}} E^{\frac{1}{2}}\,dE. \tag{81}$$

Fermi statistics (formula (30a)) gives us w for holes and electrons.

$$w_-(E) = \frac{1}{\exp\left(\frac{E-\mu}{kT}\right)+1}. \tag{82}$$

$[1-w_-(E)]$ is the probability that a level of energy E is free, i.e. that there is a hole in a level E; $w_+(E)$ is given by a formula similar to that for $w_-(E)$

$$w_+(E) = 1-w_-(E) = 1-\frac{1}{\exp\left(\frac{E-\mu}{kT}\right)+1} = \frac{1}{\exp\left(\frac{\mu-E}{kT}\right)+1}. \tag{82a}$$

Thus the probability of finding electrons in levels $(E-\mu)$ above μ equals that for finding holes $(\mu-E)$ below μ (Fig. 38). The hole energy rises as one moves downwards from the edge of the valence band; if the hole energy is reckoned downwards from μ, and the electron energy upwards from μ, we see that w_- and w_+ are equal at equal distances from μ. Equations (82) and (82a) may be simplified if $(E-\mu)$ is large in magnitude compared to kT; unity in the denominator then becomes small compared with $\exp[(E-\mu)/kT]$ and

$$w_-(E) \doteqdot \exp\left(-\frac{E-\mu}{kT}\right), \quad w_+(E) \doteqdot \exp\left(-\frac{\mu-E}{kT}\right). \tag{82b), (82c)}$$

It can be shown that μ is the chemical potential. Two conductors can be in statistical equilibrium only if their values of μ are equal; in a metal μ coincides with the Fermi level. If we reckon the electron energy upwards from μ and the hole energy downwards, we thus get a fully symmetrical pattern of energy levels, as equations (82) and (82a) show. The level filling is given by equations (82b) and (82c) for distances from μ which are large relative to kT. The probability that a

level with E equal to μ will be filled is $\frac{1}{2}$, from equation (82). Fig. 39 shows how w_- and w_+ depend on $(E-\mu)$ or $(\mu-E)$.

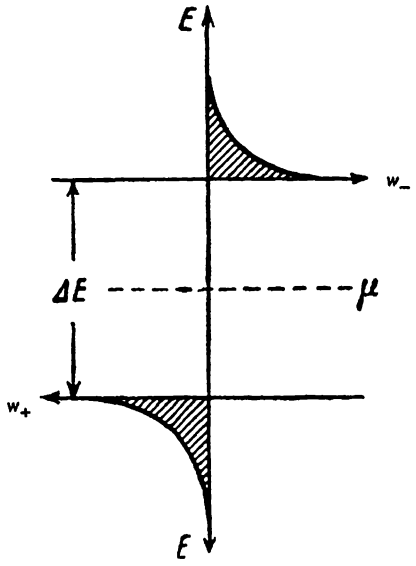


Fig. 38. The probabilities of finding electrons lying above μ by $(E-\mu)$, and of finding holes $(\mu-E)$ below μ.

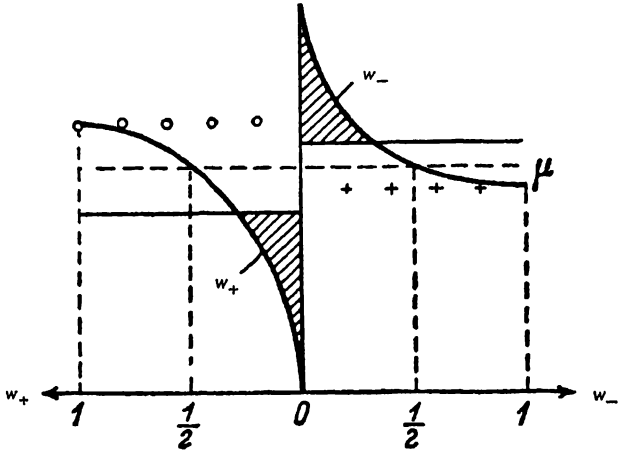


Fig. 39. Relation of w_- to $(E-\mu)$, and of w_+ to $(\mu-E)$.

The free electron and hole concentrations

We have already dealt with a semiconductor in which thermal motion has displaced some electrons from the valence to the conduction band by supplying an amount of energy ΔE_0. The number of holes in the

valence band must equal the number of electrons in the conduction band. Let the energy at the lower edge of the conduction band be E_2, and that at the top edge of the valence band be E_1. The number of levels between E and $(E+dE)$ is, as for metals, given by equation (81)

$$N(E)\,dE = 4\pi\left(\frac{2m^*}{h^2}\right)^{\frac{3}{2}}(E-E_2)^{\frac{1}{2}}\,dE.$$

Here $(E-E_2)$ is the energy in excess of that at the band edge E_2 where the kinetic energy is zero. The electrons we consider are near this edge, so the assumptions made in deriving equation (23) for metals remain true. The levels of energy E are only partially filled; the filling factor corresponds to w. In the non-degenerate case the distribution among the levels is given by

$$n(E)\,dE = w(E)\,N(E)\,dE = 4\pi\left(\frac{2m^*}{h^2}\right)^{\frac{3}{2}}(E-E_2)^{\frac{1}{2}}\exp\left(-\frac{E-\mu}{kT}\right)dE. \quad (83)$$

Let us find the total number of electrons in the conduction band at a temperature T or, in other words, the electron concentration n_-. We integrate equation (83) over all the energies in the band and get

$$n_- = \int_{E_2}^{\infty} w(E)\,N(E)\,dE = 4\pi\left(\frac{2m^*}{h^2}\right)^{\frac{3}{2}}\int_{E_2}^{\infty}(E-E_2)^{\frac{1}{2}}\exp\left(-\frac{E-\mu}{kT}\right)dE.$$

Let $(E-E_2)=\epsilon$, so $E=E_2$ means $\epsilon=0$ and $dE=d\epsilon$. Then

$$\exp\left(-\frac{E-\mu}{kT}\right) = \exp\left(-\frac{E_2-\mu}{kT}\right)\exp\left(-\frac{\epsilon}{kT}\right),$$

$$n = 4\pi\left(\frac{2m^*}{h^2}\right)^{\frac{3}{2}}\exp\left(-\frac{E_2-\mu}{kT}\right)\int_0^{\infty}\epsilon^{\frac{1}{2}}\exp\left(-\frac{\epsilon}{kT}\right)d\epsilon.$$

The latter integral is

$$\int_0^{\infty}\epsilon^{\frac{1}{2}}\exp\left(-\frac{\epsilon}{kT}\right)d\epsilon = \frac{\sqrt{\pi}}{2}(kT)^{\frac{3}{2}},$$

and so

$$n_- = 2\frac{(2\pi m_-^* kT)^{\frac{3}{2}}}{h^3}\exp\left(-\frac{E_2-\mu}{kT}\right). \quad (84)$$

The numerical values give us

$$n_- = 4{\cdot}82\times10^{15}\left(\frac{m_-^*}{m_0}T\right)^{\frac{3}{2}}\exp\left(-\frac{E_2-\mu}{kT}\right).$$

We get just the same expression for the holes in the valence band, except that the sign of $(E-\mu)$ is reversed and E_2 is replaced by E_1; m_+^* may also differ from m_-^*.

The chemical potential†

The number of electrons in the conduction band of a pure semiconductor just equals the number of empty levels or holes in the valence band, so

$$2\frac{(2\pi m_-^* kT)^{\frac{3}{2}}}{h^3}\exp\left(-\frac{E_2-\mu}{kT}\right) = 2\frac{(2\pi m_+^* kT)^{\frac{3}{2}}}{h^3}\exp\left(\frac{E_1-\mu}{kT}\right) \quad (84a)$$

or

$$(m_-^*)^{\frac{3}{2}}\exp\left(-\frac{E_2-\mu}{kT}\right) = (m_+^*)^{\frac{3}{2}}\exp\left(\frac{E_1-\mu}{kT}\right).$$

If $m_-^* = m_+^* = m^*$, then $(-E_2+\mu) = (E_1-\mu)$ and

$$\mu = \frac{E_1+E_2}{2}. \quad (85)$$

Then μ lies at the middle of the forbidden band, and

$$E_2-\mu = \mu-E_1 = \frac{\Delta E_0}{2} \quad (85a)$$

where ΔE_0 is the forbidden band width. The number of electrons in the conduction band equals the number of holes in the valence band and is here

$$n = 2\frac{(2\pi m^* kT)^{\frac{3}{2}}}{h^3}\exp\left(-\frac{\Delta E_0}{2kT}\right). \quad (86)$$

If the valence band were replaced by some number N_{eff} of electrons all of energy E_1, each such electron would need an energy ΔE_0 to reach the conduction band. Then, as with electrolytes,

$$n = N_{\text{eff}}\exp\left(-\frac{\Delta E_0}{2kT}\right). \quad (86a)$$

We can therefore take

$$N_{\text{eff}} = 2\frac{(2\pi m^* kT)^{\frac{3}{2}}}{h^3}$$

as the effective concentration of the electrons of identical energy E_1.

Equation (86) gives us the concentrations of the electrons and holes that carry the current. Taking logarithms, we have

$$\ln n = \ln 2\frac{(2\pi m^* kT)^{\frac{3}{2}}}{h^3} - \frac{\Delta E_0}{2kT} = \ln N_{\text{eff}} - \frac{\Delta E_0}{2kT}. \quad (87)$$

$\Delta E_0/2k$ is the slope of a plot of $\ln n$ against $1/T$. To obtain ΔE_0 in eV we put k in the same units

$$k = 1{\cdot}38\times10^{-16}\ \text{erg/deg} = 0{\cdot}86\times10^{-4}\ \text{eV/deg}.$$

† Editor's note. Ioffe uses the term "Fermi level" for the chemical potential only in metals. In the West it is commonly applied to semiconductors also.

We can use equation (86), which relates n to ΔE_0, instead of equation (84), which contains μ, only if μ lies at the middle of the forbidden band. If the effective masses of the electrons and holes are different, equation (84a) gives

$$(m_-^*/m_+^*)^{\frac{3}{4}} = \exp\left(\frac{E_1+E_2-2\mu}{kT}\right)$$

$$E_1+E_2-2\mu = kT\ln(m_-^*/m_+^*)^{\frac{3}{4}} \tag{88}$$

and

$$\mu = \frac{E_1+E_2}{2} - \frac{kT}{2}\ln(m_-^*/m_+^*)^{\frac{3}{4}}.$$

$(E_2-\mu)$ and $(\mu-E_1)$ depend on T, as does μ. If μ_0 is the value of μ at 0° K, we have

$$E_2-\mu = E_2-\mu_0+\frac{kT}{2}\ln(m_-^*/m_+^*)^{\frac{3}{4}}. \tag{88a}$$

Taking logarithms of equation (84) and using equation (88), we have

$$\ln n = \ln 2\frac{(2\pi m^* kT)^{\frac{3}{2}}}{h^3} - \frac{E_2-\mu_0}{kT} - \frac{3}{4}\ln\frac{m_-^*}{m_+^*}. \tag{88b}$$

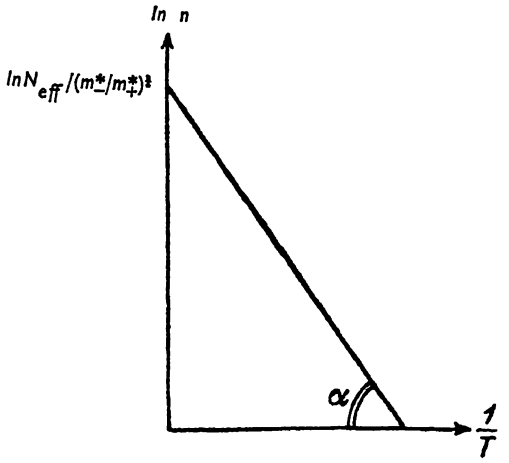


Fig. 40. Plot of $\ln n$ against $1/T$ when the holes and electrons have greatly different effective masses.

A plot of $\ln n$ against $1/T$ (Fig. 40) gives us a straight line of slope $(E_2-\mu_0)/k$, i.e. not the distance from the lower edge of the free band to μ at the temperature T used. The intercept at $1/T = 0$ also gives us, not the logarithm of the pre-exponential factor in equation (84), but the logarithm of a quantity $(m_-^*/m_+^*)^{\frac{3}{4}}$ times less.

Many semiconductors have different effective masses for their electrons and holes; equations (88a) and (88b) should therefore be considered if exact conclusions are to be drawn from the variation of σ or n with temperature.

Equation (83) defines the electrons' distribution among the levels within a band. Now $(E-E_2)^{\frac{1}{2}}$ varies only slowly with E, so the predominant factor is $\exp[-(E-\mu)kT]$. The density falls off exponentionally as E rises, just as the density ρ of a gas falls off with height h at constant T

$$\rho = \rho_0 \exp\left(-\frac{Mg(h-h_0)}{kT}\right) = \rho_0 \exp\left(-\frac{U-U_0}{kT}\right) \tag{89}$$

where Mg is the weight of one molecule of gas and U equal to Mgh is the potential energy at a height h.

The X-ray spectra of semiconductors

The soft X-ray spectra enable us to check our deductions about the distributions among the energy levels. The measurements carried out to date relate mainly to metals; only a few semiconductors have been studied, but the results agree well with the theory.

The absorption spectra define the energy needed to lift an electron from the K- or L-levels to the vacant levels; the least energy is needed to reach the lower edge of the conduction band. Above the corresponding frequency ν_0 equal to $(E_2-E_{K,L})/h$ we get absorption of strength j related to the number of absorbing electrons by $j \propto \nu^3 n(E)$. X-ray emission spectra give us the energy difference between the valence electrons and the empty K- or L-levels; these spectra end at frequencies corresponding to transitions from the upper edge of the valence band to the K- or L-levels. The maximum emission frequency is given by

$$\nu_{\max} = \frac{E_1-E_{K,L}}{h}.$$

Between $\nu_{\max}$ and ν_0 there lies a gap of width

$$\nu_0-\nu_{\max} = \frac{\Delta E_0}{h} \tag{90}$$

in which there is neither absorption nor emission. This is found, for instance, with graphite (Fig. 41). The emission intensities below $\nu_{\max}$ vary as ν^3 and give us $n(E)$ as a function of E. Many results for metals and semiconductors confirm the theory only in a general way (Fig. 15); they do not always represent exactly the electron distributions in energy. The s electrons of sodium and of other alkali metals give good agreement with theory, though the s band partially overlaps the p band in sodium. The possible transitions are then only those between the

valence band and the K-levels which correspond to the states of the innermost electrons. Only the p electrons can enter the s states. Fig. 15 does not show the electron distribution in the valence band, where there is a maximum in $n(E)$ at the band centre (Fig. 41). The distribution takes similar parabolic forms near the top and bottom edges of the band.

Fig. 41. The density of the states in the hexagonal lattice of graphite.

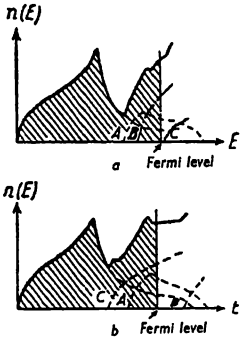


Fig. 42. The density of the states in the hexagonal lattices of (a) magnesium, (b) cadmium. A, B and C are the edges of other bands.

The situation is still more complex due to band overlap in divalent metals; the Fermi level here occurs in a region where the s and p levels overlap (Fig. 42). The shape of the curve is very different from that of Fig. 15.

Semiconductors of low resistivity

$(E_2-\mu)$ and $(\mu-E_1)$ are usually large relative to kT in semiconductors, so the factors $\exp[(E_2-\mu)/kT]$ or $\exp[(\mu-E_1)/kT]$ are very much larger than unity, as we have assumed in deducing equations (84) and (86). However, the assumption is not allowable with semiconductors of very high conductivity. The degeneracy test enables us to neglect unity compared with $\exp[(E-\mu)/kT]$ if

$$2\frac{(2\pi m^* kT)^{\frac{3}{2}}}{h^3} \gg n. \tag{91}$$

Taking m^* equal to m_0 and T equal to 300° K, we find that the degeneracy

is slight when $n \ll 2{\cdot}5 \times 10^{19}$/cc, e.g. when n is about 10^{18}/cc. Equation (84) then gives

$$n = 2\frac{(2\pi m^* kT)^{\frac{3}{2}}}{h^3} \exp\left(-\frac{E_2-\mu}{kT}\right).$$

If, however, $n \gg 2{\cdot}5 \times 10^{19}$/cc, e.g. when n is 10^{20}/cc, the electrons can be taken as degenerate, and so, as for metals near the absolute zero, equation (24a) gives

$$n = \frac{8\pi}{3}\left(\frac{2m^*}{h^2}\right)^{\frac{3}{2}} (E-E_2)^{\frac{3}{2}}.$$

The electron concentration changes gradually from that given by equation (84) to that given by equation (24a) in intermediate cases. These intermediate degeneracy cases are of great practical importance; Shifrin[5] has developed the theory of them. He takes as the degeneracy criterion the distance of μ from E_2; degeneracy for electrons begins at $(\mu-E_2)=0$; the analogous formula for holes is $(E_1-\mu)=0$. We get complete degeneracy when $(\mu-E_2) \geqslant 2kT$; if $(\mu-E_2) < -2kT$, i.e. $(E_2-\mu) > 2kT$, we can take the electrons as non-degenerate and use Boltzmann statistics.

Impurity levels

We have so far dealt with electrons in a lattice, made up of many identical atoms or molecules, in which strict stoichiometry occurs. These are ideal, not real, crystals; real crystals differ from ideal ones as follows.

1. Real semiconductors do not have the strictly periodic structures of ideal crystals; they are often polycrystalline, with portions showing cracks, or displacements and rotations relative to one another. All such faults scatter the electron waves and reduce the mobilities. However, the effects on the carrier concentrations and on the electrical parameters are slight if the faults are not numerous, as in the case of solid electrolytes. Electrons may be trapped at lattice defects, however, if the defects are such as to give discrete quantum levels.

2. Deviations from stoichiometry, i.e. excess or deficit of a component, or impurities, are much more important. These defects produce extra quantum levels distinct from those of the main lattice. The main lattice levels fuse into broad bands which contain all the valence electrons in the crystal, while the impurity levels are localized on the atoms responsible for them. These extra levels are not the same as those of the free atoms, but are not split if the impurity atoms are on average not too close together. If the impurity concentrations are such that the wave functions of the ground or excited states of the valence electrons overlap, the levels do split to give generalized impurity bands and lead to impurity conduction if the bands are not entirely full.

The conduction band in the lattice acts as an ionization region for the impurity electrons. Let the ionization potential of the free impurity atoms be P_0; the dielectric properties of the solid cause the effective sizes of the electron shells to increase by a factor ϵ, so the ionization potential is lowered by a factor ϵ^2. Now ϵ is about 10 in many semiconductors, while P_0 is 10 to 20 eV, so P_0/ϵ^2, the energy needed to transfer an impurity electron to the conduction band, is about 0·1 eV, whereas ΔE_0 (the forbidden band width) is one or more eV. The impurity levels thus fall in the forbidden band at a tenth or hundredth of an eV below the lower edge of the conduction band. The number of impurity levels is set by the number of excess atoms that act as electron donors.

Holes show a similar effect when the impurity atoms (O, S, Cl) can take up electrons from the valence band. The energy required for the transfer is again reduced by a factor ϵ^2. These acceptor levels lie a distance ΔE above the top of the valence band (ΔE is tenths or hundredths of an eV).

Pairs of electrons from adjacent atoms form covalent bonds along the lines joining the atoms in covalent lattices. In diamond, silicon and germanium crystals every atom is bound to four others by its four valence electrons; only four of the valence electrons are used in the bonds, if the quadrivalent atom is replaced by one of higher valency, and the remainder, much less strongly bound, may pass to the conduction band. The impurity atoms become electron donors, and the semiconductor shows electronic conduction. If the foreign atom is one of valency less than four, its valence electrons are too few for completion of the bonds to the adjacent atoms; thermal agitation may bring up the lacking electrons from the valence band and thus leave behind the same number of holes. The impurity acts as an acceptor and gives hole conduction.

ΔE_0 is about 1·5 eV in cuprous oxide, while ΔE is about 0·6 eV; in this case 0·01% of impurity atoms suffices to change the hole content appreciably at room temperature. In pure cuprous oxide $(\mu - E_1) = 0{\cdot}75$ eV (if we take $m^* = m_0$), while $(\mu - E_1) = 0{\cdot}3$ eV if the chemical potential level lies half way between the valence band and the impurity levels. In the second case $\exp[-(\mu - E_1)/kT]$ is 3×10^6 times larger than in the first, so, although the impurity levels are 10^4 times fewer than the main lattice levels, the impurities give a hole concentration 100 times larger than is found in pure Cu_2O.

These results relate to room temperature; the ratio between the holes from the two sources changes at high temperatures. At 600° C we have $\exp[-(\mu - E_1)/kT]$ for the impurities only 1000 times larger than for the main lattice; the extra hole concentration due to the

0·01% impurity is thus less than that which would be present in the pure compound.

Zhuze and Kurchatov[6] have studied the conductivities of many cuprous oxide samples with various excess oxygen contents over the range from $-180°$ C to 900° C. In all cases the high-temperature conductivity was given by

$$\sigma = \sigma_0 \exp\left(-\frac{1{\cdot}44}{2kT}\right).$$

This corresponds to intrinsic excitation across the energy gap ΔE_0 equal to 1·44 eV. At lower temperatures it was found that

$$\sigma = \sigma_1 \exp\left(-\frac{\Delta E}{2kT}\right)$$

where $\Delta E = 0{\cdot}3$ to 0·6 eV; this corresponds to holes freed by the excess oxygen. At intermediate temperatures

$$\sigma = \sigma_0 \exp\left(-\frac{1{\cdot}44}{2kT}\right) + \sigma_1 \exp\left(-\frac{0{\cdot}6}{2kT}\right).$$

The greater the amount of excess oxygen the larger σ, and the higher the temperature at which impurity conduction prevailed. $\Delta E/2$ was about 0·3 eV at all temperatures in all specimens, though it fell somewhat as the impurity content rose. The impurity levels have a major effect on μ.

The sum of the positive and negative charges in a crystal must be zero. Let us suppose that there are N impurity levels at a distance ΔE from the band edge such that almost all the free charges originate from these levels (as is the case with Cu_2O at room temperature). The minute number of electrons transferred from the valence band into the conduction band can then be ignored; only those that come from the impurity atoms need be considered. The impurity atoms that have lost electrons become positive ions. We get negative ions and holes when electrons pass from the valence band to the impurity atoms. In both cases the free charge concentrations are given by

$$n = \sqrt{(NN_{\text{eff}})} \exp\left(-\frac{\Delta E}{2kT}\right) = \sqrt{2}\sqrt{N}\left(\frac{2\pi m^* kT}{h^2}\right)^{\frac{3}{4}} \exp\left(-\frac{\Delta E}{2kT}\right); \quad (92)$$

μ is then given by $$E_2 - \mu = \frac{\Delta E}{2} - \frac{1}{2}kT \ln \frac{N}{N_{\text{eff}}}. \quad (92a)$$

The second term on the right is usually small compared to the first, so μ lies not near the middle of the forbidden band (Fig. 43a), but at the centre of the gap ΔE between the impurity levels and the band edge (Fig. 43, b and c).

Acceptor impurities can be considered in the same way as donors (Fig. 43c); μ shifts from a position such that $(\mu - E_1) \doteqdot \Delta E_0/2$ to one where $(\mu - E_1) \doteqdot \Delta E/2$. In Cu_2O at high temperatures, where the main lattice gives the free carriers, μ lies at a level given by $(\mu - E) = \Delta E_0/2$, i.e. at 0·75 eV above the top edge of the valence band. At low temperatures the excess oxygen gives most of the holes, and $(\mu - E_1) = \Delta E/2 = 0{\cdot}3$ eV. At intermediate temperatures the two sources of holes are of comparable significance; μ then moves in such a way that the total numbers of positive and negative charges are equal.

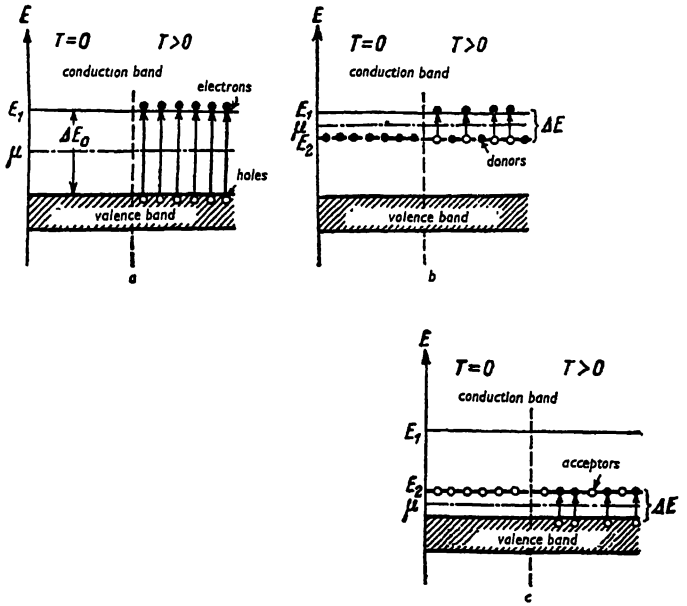


Fig. 43. The energy schemes at $T = 0$ and $T > 0$ for (a) a semiconductor entirely free from impurities, (b) a semiconductor with a donor impurity, and (c) a semiconductor with an acceptor impurity. μ is the level of the chemical potential.

We have seen previously that μ varies with temperature even in a pure substance if the holes and electrons do not have equal effective masses, as is usually the case. The position of the chemical potential relative to the band edges determines the free carrier concentrations.

Semiconductors may contain donors and acceptors at the same time; if the donor levels lie above the acceptor levels the electrons pass from the donors to the acceptors, making the atoms positively and negatively charged respectively. Whichever type of atom exists in smaller numbers will be present entirely as ions, while the type that predominates will be present partly as ions and partly as neutral atoms; the number of the

latter will be equal to the difference between the numbers of donors and acceptors. This difference alone determines the conductivity type, which will be either electron or hole, depending on whether donors or acceptors respectively are in the majority. The chemical potential will lie at a corresponding level. If the donors and acceptors are equal in numbers (Fig. 44) and if their levels are equidistant from the corresponding band edges, and if, moreover, $m_+^* = m_-^*$, then μ will lie just half-way between the band edges; $(E_2 - \mu) = (\mu - E_1) = \Delta E_0/2$, just as in a perfectly pure crystal. The electron and hole concentrations will then be the same as in the pure substances

$$n_+ = n_- = 2\frac{(2\pi m^* kT)^{\frac{3}{2}}}{h^3}\exp\left(-\frac{\Delta E_0}{2kT}\right) = N_{\text{eff}}\exp\left(-\frac{\Delta E_0}{2kT}\right) \qquad (93)$$

and the concentrations will vary with temperature in the same way as for the pure substance.

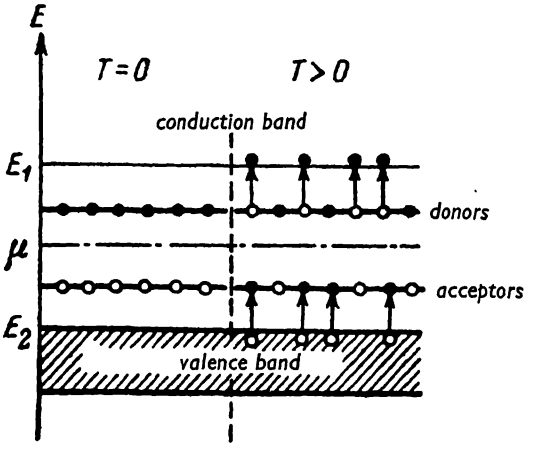


Fig. 44. The energy scheme of an impurity semiconductor with equal numbers of donors and acceptors.

The impurities alter the mobilities and not only the free carrier concentrations; the values are reduced below those for the pure substance. The type of scattering also changes. The electrons are more mobile if they are scattered only by thermal fluctuations; the ionized impurities have the effect of lowering the mobility.

Equation (92) applies only if N is large compared with the free electron concentration n; n, of course, rises more slowly as it nears N, and becomes constant when all the impurity atoms have lost or gained electrons. As n nears N, e.g. as T rises, the donors become depleted and n tends to its limiting value N; μ would then fall towards $-\infty$ if

the donor levels were the sole source of free electrons. However, the main lattice begins to give electrons before this happens, so μ falls only to the level it has in the pure substance; any further rise in the number of free carriers will be due to the main lattice. Fig. 45 shows how the conductivities of some samples of silicon vary with temperature. The depletion region is here seen to correspond to a fall in σ as the temperature rises, because the mobility falls (curves 2 to 4).

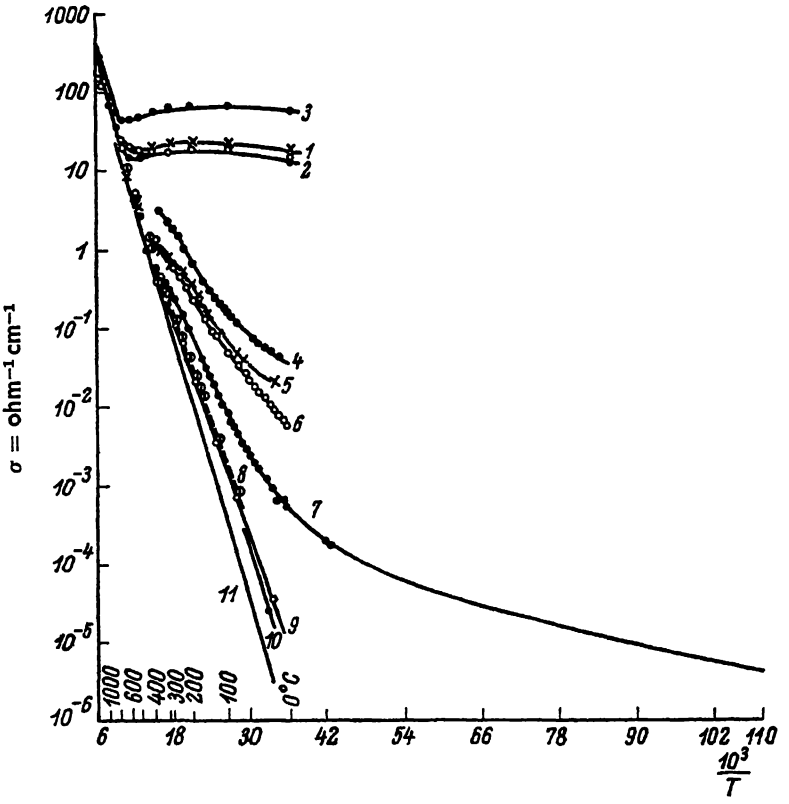


Fig. 45. The electrical conductivity of silicon specimens with various impurity contents as a function of temperature.

μ must lie at the same level in metals and semiconductors which are in equilibrium with one another. The position of μ in a semiconductor is dealt with above; in metals μ lies at the limit of the Fermi distribution.

Surface states

Tamm[7] has shown that extra levels are found at crystal surfaces, where the periodic lattice potential goes over to the fixed external

potential. Every surface atom produces one ground and one excited level (Fig. 46); electrons from the valence band enter and fill only half of these levels, just as the $3s^1$ electrons in sodium fill only half the band; that is why sodium is a metal. We would thus expect that surface levels should give the surface layer metallic properties. However, no experimental proof of this exists, possibly because any type of surface flaws, or adsorbed ions or molecules, may cause the levels in the separate atoms to diverge, and thus prevent free exchange of electrons between them.

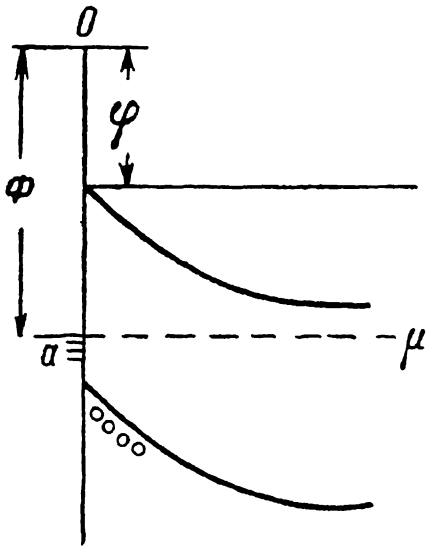


Fig. 46. The energy scheme of a semiconductor with surface levels (a). ϕ is their work function and ψ is the external work function.

The disturbances due to the free surface persist within the crystal but die away rapidly. Taking the mean atomic spacing as 3×10^{-8} cm, we have 10^{15} atoms per cm^2, and thus 2×10^{15} extra levels; the main lattice has 4×10^{22} levels per cm^3, i.e. 2×10^7 times more. Even the impurity levels (10^{19} per cm^3 at 0·1% impurity) are much more numerous than the surface ones. Tamm's surface levels thus cause no major changes in the bulk properties, but effects which occur in the thin surface layer may be decisively affected. The contact resistance against a metal or another semiconductor, rectification, and high-frequency amplification, are all determined by layers 10^{-4} to 10^{-6} cm thick. Here the main lattice levels vary in number from 4×10^{18} to 4×10^{16}/cm^2, and the impurity levels (0·1% impurity) from 4×10^{15} to 4×10^{13}/cm^2. The 2×10^{15}/cm^2 surface levels in the forbidden band may then alter

the effects entirely. The surface levels are, in fact, used to explain the effects found under these conditions.

Conditions similar to those found at external surfaces also exist in cracks and, to a lesser extent, in lattice distortions. We may here again expect local discrete levels within the forbidden band.

Polarons

We have so far dealt with the energy levels of the unexcited lattice or of the impurity atoms in it; we have not allowed for the lattice changes caused by the electrons or holes freed by light or by thermal causes, and have not dealt with the way in which the electrons are freed. The electron and its parent (ionized) atom polarize their surroundings, the more so the higher the dielectric constant. There are two types of polarization: (1) Electron displacements, which take some 10^{-15} sec, i.e. times of the order corresponding to optical frequencies; here ϵ is close to the square of n, the refractive index at long wavelengths; and (2) Lattice ion displacements (e.g. in alkali halides), or rotations of molecular dipoles, or, sometimes, reorientations of entire regions (domains). These latter displacements take much longer to appear (10^{-13} sec or more), and $\epsilon \geqslant n^2$; ϵ is often much larger than in case 1). The polarization always reduces the potential energy; the energy levels of the electrons are lowered. If some such level before polarization lay at the lower edge of the conduction band, it would enter the forbidden band as the lattice arrangement changed. The fall could be of the order of 0·1 eV in ionic crystals. The electron takes its state of polarization with it as it moves; new regions are polarized while others revert to their previous states. Pekar[8] has studied these free charges moving in a polarized medium and has called them polarons. The polaron states form a continuous band in which the polarons move, just as do electrons in the conduction band, and the mobilities are of the same order. The only requirement is that there should be time for the medium to be polarized.

Let U be the fall in the energy level of an electron caused by a polaron. The probability that the polaron levels will be filled is

$$\exp\left(-\frac{E-\mu-U}{kT}\right),$$

while the probability that the levels near the edge of the conduction band will be filled is

$$\exp\left(-\frac{E-\mu}{kT}\right), \tag{93a}$$

i.e. $\exp(U/kT)$ times smaller. $\mathrm{Exp}(U/kT)$ is about 10,000 if U is 0·2 eV (as in some polar crystals) and kT is 0·03 eV.

Excitons

Frenkel'[9] has shown that special states related to sets of electrons, which he calls excitons, can arise in crystals. An electron is transferred to an excited state by a photon (or perhaps by thermal means); the resulting hole and electron behave as a single system, similar to a hydrogen atom; less energy is needed than is required to form a free electron plus a free hole. The electron remains bound to the hole and the two may move together from one atom to another of the same properties. The exciton diffuses at random until it encounters conditions which cause one of the charges to become bound, or the hole and electron to combine. In the first case the other charge becomes free and increases the carrier concentration, and hence the conductivity. In the second case the energy is released as heat or light.

The exciton pair carries no current; excitons produced by light absorption do not of themselves lead directly to photoconduction. The lattice absorbs the light energy only when the exciton vanishes. This is, according to Frenkel', how light is absorbed without a photoelectric effect; the latter occurs only after one of the charges has been trapped at a localized impurity level. Zhuze and Ryvkin[10] have used the exciton to explain why the strong fundamental absorption in cuprous oxide gives but a small photoconductivity. The absorbed light forms excitons, some of which later meet impurity levels, are destroyed, and give their energy to impurity electrons. The holes left behind impart the appropriate type of conductivity to the cuprous oxide. The probability that a hole will be freed is thus proportional to the number of impurity levels, whereas the absorption is determined by the pure cuprous oxide.

Gross and Karryev[11] have observed exciton spectra in absorption by cuprous oxide and some other semiconductors. These spectra consist of sets of sharp hydrogen-like lines which merge into a continuous ionization band towards the higher frequencies. Excitons differ from hydrogen atoms in that the positive and negative carriers are of about the same mass.

The energy levels of the electron in the Coulomb field of the positive charge lie somewhat below the edge of the conduction band. The electron needs energy to transfer it from the exciton to the free state, where it can carry current and give a photoelectric effect. Thermal sources can supply this energy. This may explain why the photocurrents are so low in alkali halide crystals at low temperatures (the thermal energies are too low to ionize the excitons) and rise so rapidly with temperature up to a certain limit; kT is larger than the distance of the exciton from the edge of the conduction band above the limit. If this is in fact true, this distance should be given by the temperature

that corresponds to the sharp rise in photoconductivity. For KCl this is 200° K, where kT is about 0·02 eV.

In cuprous oxide the distance of any exciton state from the conduction band edge is defined by the energy difference between the appropriate absorption line and the start of the continuum. Gross found values of 0·1 to 0·2 eV; he found similar effects in some other semiconductors, mainly in those which show high photoconductivities.

The energy levels of semiconductors

We can now give a graphic scheme for the energy levels of a semiconductor. Let us first consider the periodic potential in which the valence electrons of a crystal move. The *K*-, *L*- and *M*-group electrons are strongly bound to the nuclei; they interact so weakly with nearby atoms that the levels are not split appreciably, so they remain almost as sharp as in the free atoms. The wave functions of the valence electrons overlap with those of nearby atoms and give a single system (a band) with $2N$ states. The more strongly the valence electrons are bound, and the weaker their interactions with nearby atoms, the narrower the band formed from a discrete level in the free atom.

The positive ions can be taken as placed at the lattice nodes, with the space between filled with an electron continuum, in a neutral crystal; this representation can be used to show the valence electron levels. The potential energy of an electron falls sharply near the positive ions and is largest between the ions. The maxima may be higher than the quantum energies of the electrons, but the energy barrier so formed is easily overcome if it is no larger than 1 to 3 eV, if the atomic spacing is small. Resonant electrons in adjacent atoms exchange places through these barriers some 10^{10} to 10^{15} times per second.

Fig. 47 shows the energy levels and distributions for various types of semiconductor, while Fig. 48 shows the exciton and polaron levels. These diagrams all show the quantum state energies of electrons in a crystal; as in a free atom, the states are defined by the complete sets of quantum numbers; the corresponding effects appear in the selection rules, for instance. The X-ray spectra of semiconductors and metals do, in fact, on close analysis show effects related to quantized parameters other than the energy.

The selection rules complicate the relations between the optical absorption spectra and the photoelectric effect, on the one hand, and the forbidden band width, on the other. The rules require that the wave numbers k must be the same for any electron transition in a semiconductor. It may occur that optical transitions from levels at the top edge of the valence band to the lower edge of the conduction band are forbidden, and the only allowed transitions are those to some other

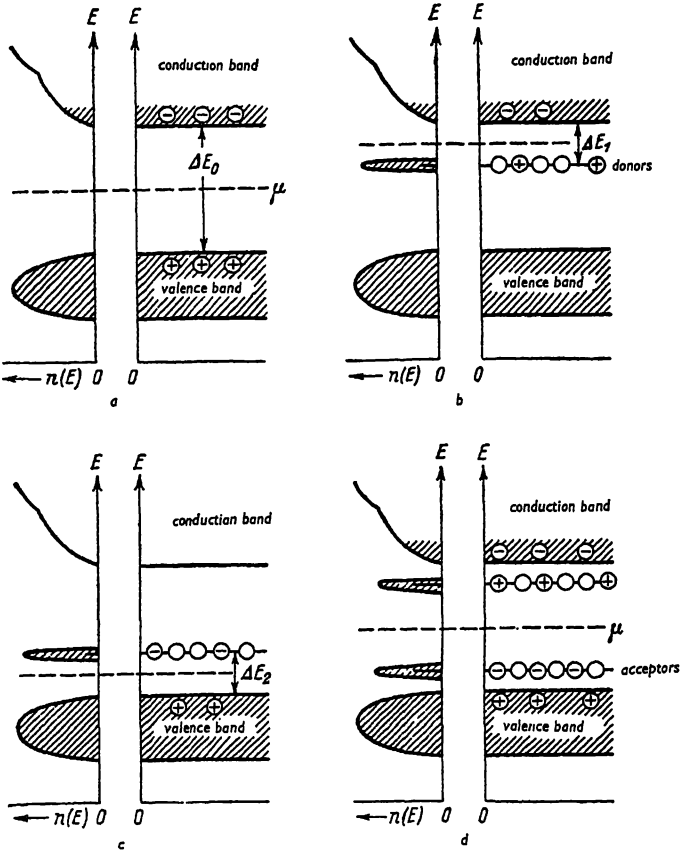


Fig. 47. The energy schemes of a pure semiconductor and of impurity semiconductors at $T>0$: (a) Entirely pure semiconductor, (b) a semiconductor with a donor impurity, (c) a semiconductor with an acceptor impurity, (d) a semiconductor with equal numbers of donors and acceptors. $n(E)$ is the number of quantized states per unit energy per unit volume.

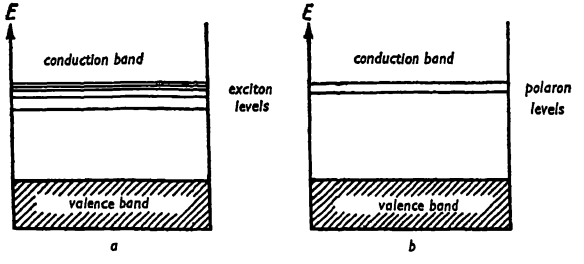


Fig. 48. The energy schemes of semiconductors with (a) exciton and (b) polaron levels.

system of levels which does not include this lower edge. The wave numbers may be of different orders as we move away from the edges of the upper and lower bands. The only transitions allowed from the top of the valence band may be to free levels far from the edge of the conduction band, while electrons may be able to enter the lower edge of the conduction band from deep-lying valence band levels only. The wave numbers need not be equal for electrons coming from lattice defects, since the lattice itself here takes part in the transition.

The optical rules do not apply to thermal transitions, in which heavy atoms and phonons of momentum $h\nu/v$ are active (v is the speed of sound); photons of very low momentum $h\nu/c$ (c is the speed of light) are alone active in optical transitions. The minimum thermal energy needed to transfer an electron from the valence to the conduction band is thus always just equal to the forbidden band width, which is not necessarily true of optical transitions.

Another type of impurity conduction, which does not require electrons to enter the conduction band, is possible and sometimes occurs; an impurity band is formed when the concentration of impurities is so high that their wave functions overlap. If the lattice constant is 3 Å and the impurity content 2% the mean distance D between impurity atoms is 10^{-7} cm. The diameter d of the orbit of a singly charged ion in a medium with $\epsilon = 10$ is $d_0\epsilon$, where d_0 is the size of a hydrogen atom orbit (1·06 Å); then D is close to d.

Excited states make it even easier for the wave functions to overlap; 0·5% impurity is enough to cause overlap for excited states with the quantum number $n = 2$. If all the N impurity atoms are in the same state the system of excited levels will form a free band to which electrons may rise from the valence band while absorbing less energy than is needed to reach the conduction band of the main lattice. Even if the wave functions do not overlap to give a new band, the impurity atoms may interact in such a way that their electron states come close to the edges of the conduction or valence bands. This is probably the origin of the empirical Meyer–Neldel law relating the dissociation energy of an impurity to its concentration, via the relation between the two constants σ_0 and ΔE which define σ for a semiconductor

$$\sigma = \sigma_0 \exp\left(-\frac{\Delta E}{2kT}\right).$$

Highly conducting semiconductors with high impurity concentrations give

$$\Delta E = \Delta E(0) - \alpha \ln \sigma_0, \tag{94}$$

i.e. the ionization energy drops as the impurity concentration (which

determines σ_0) rises; semiconductors with low concentrations and $\sigma < 10^{-2}$ ohm^{-1} cm^{-1} give

$$\Delta E = \Delta E(0) + \beta \ln \sigma_0 \tag{94a}$$

at room temperature. Here, where the impurity atoms interact only negligibly, the work of ionization rises with the impurity concentration.

Busch has put forward an explanation for equation (94a). The concentration of dissolved impurities N is equivalent to that in equilibrium at some high temperature T_k; on cooling this concentration N is frozen in, being given by

$$N = A \exp\left(-\frac{E_k}{kT_k}\right)$$

where E_k is the energy needed for an impurity atom to enter the solid solution. It is supposed that ΔE and the impurity concentration, which determines σ_0, both increase as T_k increases.

This quenching effect has been observed in semiconductors with hole (Cu_2O) and electron (ZnO) conduction. The excess oxygen in the Cu_2O and the partial pressure of oxygen in the surrounding medium reach equilibrium rapidly above 800° C; the time needed is very long below 800° C. Rapidly cooled Cu_2O retains its excess oxygen.

Another conduction mechanism occurs in some cases, especially in oxides. Semiconductors containing atoms with more than one possible valency (e.g. Fe, Mn) are sometimes crystalline and built up from alternating di- and trivalent atoms so spaced that electrons can be exchanged freely. The net energy change is zero when an electron passes from a divalent atom to a trivalent one; the electron does not need to enter the conduction band. However, the exchange may need an activation energy u, and thermal fluctuations must provide this. The transition probability will be determined by a factor $\exp(-u/kT)$, where, however, u does not equal ΔE. This mechanism is found in an extensive class of semiconductors known as the spinels (inverted spinels), which are dealt with in the last chapter.

The present state of semiconductor theory

The quantum theory of semiconductors has some serious major flaws, although it has been used with great success in solid-state physics. The results it gives, while qualitatively true are not quantitatively correct, and it often does not give precise numerical predictions; they are, thus, often of no practical use.

The band theory is based on the periodic nature of lattices, which causes the wave functions to be periodic, and hence determines the laws of motion of electron and holes. The same band structures are found in amorphous and liquid semiconductors, which lack long-range order;

the main properties of these substances are determined by nearest-neighbour interactions. The spacings and arrays formed by the atoms are more responsible for the properties than is the periodic character of the lattice.

The free motion of an electron in a periodic field of any spacing, no matter what the height of the barriers between the atoms, is in practice limited by the necessity for the wave functions to overlap and by the time required for the transitions. Band theory considers the behaviour of an additional electron injected into the self-consistent field produced by the ions in their instantaneous positions and by the related electrons. The direct interactions with other electrons are neglected. It is also arbitrarily assumed that the processes are adiabatic; approximations are made, as in the case of the effective masses. There are thus major objections to the one-electron band theory.

These defects in the theory mean that it cannot predict any of the constants of a semiconductor exactly; sometimes even qualitative predictions about relationships to temperature, pressure, magnetic field, etc., are not verified by experiment. Thus the predicted variation of the mobility u with T is found to agree roughly with theory only for bodies with purely homopolar bonds. In polar crystals, for which the theory gives $u \propto T^{-\frac{1}{2}}$ above the Debye temperature, experiment normally gives a very much more rapid fall in u, even up to the highest temperatures. In some cases $u \propto T^{-2\cdot5}$; this variation has, so far, not been explained.

The mean free path, and its relation to the kinetic energy, become meaningless for ionic semiconductors, since the mean free paths are less than the interatomic distances. The electron wavelength here exceeds the sizes of the scattering centres, while the indeterminacy in the kinetic energy is much larger than the kinetic energy itself. The mechanism whereby the energy taken up from the electric field is transferred to the lattice, i.e. the mechanism of Joule heating, can hardly be represented as a steady build-up in the energy during a free path followed by transfer of energy $h\nu_0$ as a phonon of frequency ν_0. The resistance to motion which the electron encounters must be explained in another way.

Theoretical predictions of how α (the thermoelectric power) varies with T agree with experiment in only a few cases, for limited temperature ranges. Most semiconductors show thermoelectric powers diverging widely from the theoretical values; no reason has yet been found for this. The prediction that $\alpha \propto 1/T$ would, if extrapolated to 0° K, imply an infinite value for α, whereas we expect α to tend to zero as $T \rightarrow 0$, as is the case for metals with degenerate electron systems.

The theory predicts the wrong magnetoresistance effects; only semiconductors with atomic lattices and pure covalent bonds agree well

with theory. The predictions are found to apply only partially to other semiconductors.

Some of the theoretical predictions have not been checked against many materials; very few data exist on effective masses, on mean free paths, on the contributions of electrons and holes to mixed conduction, on the properties of semiconductors at low temperatures, on the polaron conduction mechanism, on carrier lifetimes, on recombination mechanisms and on many other features of semiconductors.

We must thus recognize that our theoretical and practical studies of semiconductors, although numerous, are as yet inadequate. There are many unexplained contradictions and frequently no certain correlations between theory and experiment. The modern quantum theory is nevertheless reliable, even if inadequate to give a full and correct understanding of the enormous variety of effects found in semiconductors and of technological use. Although there are many problems which have not been cleared up, the theory can explain most of the more important effects qualitatively, if not quantitatively.

We shall next deal with the laws governing the main effects found in semiconductors, with the special properties of boundary layers, and with methods of measurement. The last chapter describes various semiconductors and the effects which they show.

We shall review the experimental data, the considerable advances which have already been made, the contradictions as yet unresolved, and the major demands on semiconductor physics made by industrial requirements.

References

GENERAL

F. Seitz, *The Modern Theory of Solids.* New York-London, 1940.

B. I. Davydov and I. M. Shmushkevich, "Theory of electronic semiconductors—I", *Usp. Fiz. Nauk*, **24**, No. 1 (1940).

SPECIFIC PROBLEMS

1. E. I. Adirovich, *Nekotorye voprosy teorii lyuminestsentsii kristallov* (*Some Problems of the Theory of Luminescence of Crystals*). Moscow-Leningrad, Gos. Izd. Tekh.-Teor. Lit., 1951.
2. S. I. Pekar, "The theory of F-centres", *Zhur. Eksper. Teor. Fiz.*, **20**, 510 (1950).
3. S. I. Pekar, "The theory of recombination of electrons in semiconductors", *Zhur. Eksper. Teor. Fiz.*, **20**, 267 (1950).
4. L. Sosnovskii, "Recombination in current carrier collisions in semiconductors", *Izv. Akad. Nauk SSSR, Ser. Fiz.*, **21**, 70 (1957).
5. K. S. Shifrin, "On the theory of electrical properties of semi-metals", *Zhur. Tekh. Fiz.*, **14**, 1, 49 (1944).
6. V. P. Zhuze and B. V. Kurchatov, "On the electrical conductivity of cuprous oxide", *Zhur. Eksper. Teor. Fiz.*, **2**, 5, 309 (1932).

7. I. E. Tamm, "A feasible model of electron binding at crystal faces", *Phys. Zs. S.U.*, **1**, 733 (1932).
8. S. I. Pekar, *Issledovaniya po elektronnoi teorii kristallov* (*Studies in the Field of the Electronic Theory of Crystals*). Moscow-Leningrad, Gos. Izd. Tekh.-Teor. Lit., 1951.
9. Ya. I. Frenkel', "On the transformation of light into heat in solids—I and II", *Phys. Rev.*, **37**, 17, 1276 (1931).
10. V. P. Zhuze and S. M. Ryvkin, "Exciton nature of light absorption and impurity photoconduction", *Dokl. Akad. Nauk SSSR*, **77**, 241 (1951).
11. E. F. Gross and N. A. Karryev, "Optical spectrum of an exciton", *Dokl. Akad. Nauk SSSR*, **84**, 471 (1951).

CHAPTER

5

Physical Phenomena in Semiconductors

Properties of the crystal lattice

Short- and long-range order

The quantum theory of metals and semiconductors regards these materials as made up of a system of positive ions, formed by the nuclei and the localized electrons belonging to the inner shells K, L, M..., together with the valence electrons which may be present in a free band.

The large difference between the masses of an ion and an electron makes it possible to consider the electronic processes as taking place in the field of ions which are effectively stationary. The displacements of ions due to thermal agitation or valence forces take place with speeds which are about a thousand times less than the velocities of electrons. The thermal velocities of electrons are of the order of 10^7 cm/sec and their mean free paths are less than 10^{-6} cm. Consequently, the mean free time between collisions for an electron does not exceed 10^{-13} sec, while the thermal vibrations of the ions have periods of 10^{-11} to 10^{-12} sec. The displacements of electrons within the boundaries of an atom take place in times which are even shorter than the mean free time between collisions, of the order of 10^{-15} sec.

These facts allow one to consider the displacements of ions and slow processes in a crystal lattice independently from the electronic phenomena. The disposition of atoms in a solid may be considered from the point of view of short- and long-range order.

By short-range order we understand the relative disposition around an atom of its nearest neighbours. The inter-atomic distance and the mutual disposition of atoms determine the magnitude and direction of the forces of interaction and also the overlap of the wave functions. The short-range order determines, therefore, the structure of energy levels and the width of the forbidden gap and, consequently, the concentrations of free electrons and holes. In fact, Regel demonstrated experimentally the validity of this reasoning. Some semiconductors retain this type of conduction after melting and the same density of electrons as in the solid state.[1]

In a crystal the short-range order is determined by the unit cell. X-ray analysis shows that in amorphous materials each atom is surrounded by neighbours at approximately the same distances and in the same directions as in the unit cell of a crystal of the same material (Fig. 49). However, small irregularities in individual cells are randomly disposed with the result that cells separated by a few lattice distances are differently oriented.

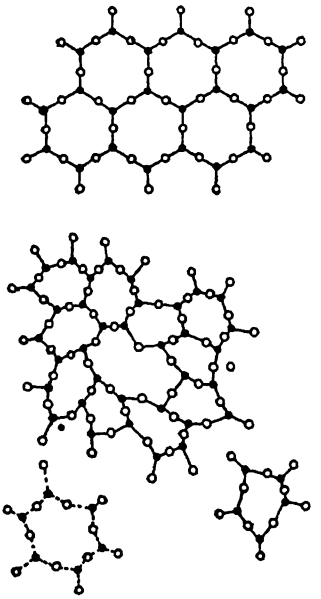


Fig. 49. Crystalline and amorphous forms of quartz.

Long-range order is found in crystals and is characterized by the fact that any structural element of the lattice (a given atom, say) is found in any particular direction at regular intervals. This implies that if the given element is found at a distance a, there is a very high probability of finding it at intervals $2a, 3a, \ldots, 1000a$ and, conversely, the probability of finding this element in an intermediate position is extremely small.

A given order seldom extends over large regions of a single crystal; strict regularity usually obtains over distances not larger than 1000 atomic spacings. The imperfect conditions of growth, and mechanical stresses, are known to give rise to defects in the structure of real crystals. Even this degree of regularity is sufficient, however, for the manifestation of all phenomena, associated with the interference of X-rays and electronic waves, which are characteristic of the crystal. Furthermore,

all quantum mechanical deductions, based on the periodicity of the crystal lattice for electrons, are valid.

Thermal agitation gives rise to slight displacements of atoms from their equilibrium positions but it does not destroy the long-range order. The implication is that there is still a much higher probability of finding an atom at or near an interval na where n is an integer, than at intermediate positions. In fact it is found that the sharpness of interference phenomena is not reduced by thermal agitation. It does depend strongly, on the other hand, upon structural inhomogeneities. Interference lines become blurred if the long-range order extends over less than ten inter-atomic distances while, in the case of order extending over only two or three lattice constants, the interference maxima are barely discernible.

The electronic wavelength in a semiconductor at room temperature amounts to approximately 7×10^{-7} cm, i.e. 15–20 inter-atomic distances. The mean free path in a good crystal exceeds 10^{-6} cm and attains the order of a hundred inter-atomic distances. An electron travels with the minimum of scattering only in the case where full long-range order is maintained over distances of this free path length.

Thermal agitation has considerable influence on the motion of electrons even though it does not disturb the long-range order. It gives rise to localized fluctuations resulting in compressions and dilatations of the atomic structure and, thus, disturbs the perfection of the lattice. Such departures from periodicity extending over distances, of the order of 10^{-7} cm, equal to a quarter of the electronic wavelength result in particularly strong electron scattering. Each deflection through a large angle must satisfy quantum mechanical conditions; the wave vector may change by $2\pi/a$, with the energy $h\nu$ given to, or taken from, the thermal energy of the lattice. The process of scattering of electronic waves by the waves of thermal agitation of the lattice may be described in terms of collisions between electrons and phonons, with exchange of their quasi-momenta and energies.

Electrons are similarly scattered by other irregularities which disturb the periodicity of the lattice over distances not less than $\lambda/4$, where λ is the electronic wavelength. Grain boundaries, dislocations due to mechanical stresses, foreign inclusions and impurities, and lattice defects all contribute to the additional scattering of electrons. Especially effective in this respect are ions; their charges deflect the paths of slow electrons more strongly than those of faster ones.

Long-range order is thus an indispensable condition for the free propagation of electrons. A disturbance in this order, over distances of 10^{-7} cm, results in scattering and a reduction of the mean free path. A large mean free path and a high mobility are attainable only in

materials exhibiting long-range order. The free path length is reduced by thermal agitation and lattice imperfections. Supposing that, at the melting point, the mean free path has already reached the minimum value typical of the amorphous state, then no appreciable change of either the free carrier concentration or the mobility is to be expected upon melting. Melting is, however, always accompanied by a change of density and, sometimes, also in the disposition of neighbouring atoms. As a consequence, some discontinuous change of conductivity upon melting is always to be found.

Semiconductor crystal lattices

The properties of semiconductors depend, in the first instance, upon the forces binding the atoms into a continuous solid body. Three types of forces may be distinguished: valence, ionic and Van der Waals type forces. The hydrogen and metallic bonds occupy a special place.

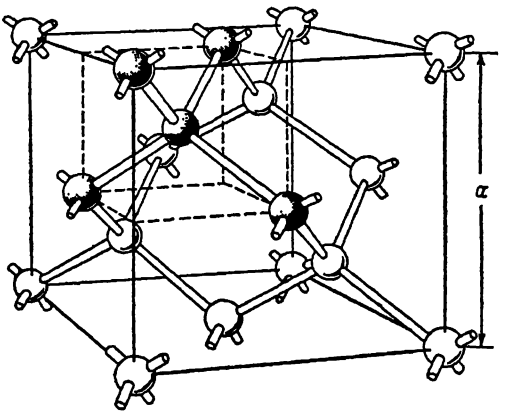


Fig. 50. Crystal structure of diamond. The diagram shows the tetrahedral bonds between atoms in the lattice.

Valence forces bind together atoms in molecules, as well as in crystals, by means of the sharing of one or two pairs of electrons. Valence forces result in extremely powerful bonds, giving rise to high melting points, high latent heats of vaporization and considerable hardness. In a crystal these bonds are oriented in definite crystallographic directions. Examples of this may be seen in diamond and silicon.

In diamond each tetravalent carbon atom is bound to its four nearest neighbours, which are identical atoms disposed tetrahedrally around it (Fig. 50). One electron from each atom participates in each of the four directional bonds. Valence bonds differ from ionic and

metallic bonds in that the pair of electrons binding any two atoms together does not participate, at the same time, in any other bonds in the crystal. Such bonds, due to quantum exchange forces, are called saturated bonds.

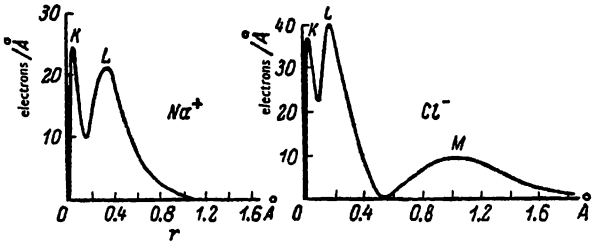


Fig. 51. The distribution of negative charge in the electronic clouds of Na^+ and Cl^-.

Fig. 52. The electronic clouds for: (a) Van der Waals type forces; (b) ionic bonds; (c) covalent bonds.

Ionic bonds, too, are formed by electrons but, in contrast to the valence bond, an electron is transferred from the more electro-positive atom to the more electro-negative atom, thus leaving the former as a positive ion and the latter as a negative ion. The distribution of electronic charge in the ions Na^+ and Cl^- is shown in Fig. 51. As the atoms approach each other, the transfer of an electron reduces their overall energy. Two ions with opposite charges, attracting each other with electrostatic forces, form a molecule. In a crystal, positive and negative ions are disposed in alternate positions; the attraction of positive ions to their negative neighbours binds the totality of ions into a single continuous lattice. Ionic binding forces are not saturated; a system may comprise any number of ions. Examples of crystals with ionic binding are given by NaCl, LiF, $CaCO_3$. In such crystals one cannot distinguish individual molecules.

The differences between the distributions of electronic charge are represented figuratively in Fig. 52 for separate molecules while Fig. 53 shows the distribution of charge in a crystal by means of equipotential lines.

It is evident that there exist many intermediate cases between the two extremes of ionic bonds, with the complete transfer of electronic

charge from one atom to another, and valence bonds, where the electronic distribution is strictly symmetrical. Molecules exhibiting electric dipoles are similar to ions and may be bound into a solid or liquid body by means of electrostatic forces between the dipoles.

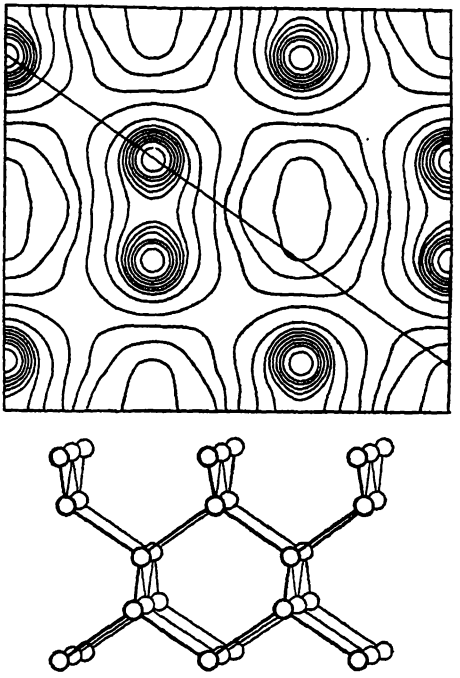


Fig. 53. The distribution of electron density and the lattice model for diamond.

The third type of force is represented by the bond between electrically neutral atoms, in which dipoles are periodically induced by mutual polarization. Binding of this type is much weaker than in the former two cases; the binding energy decreases as the sixth power of the interatomic distance. These forces are found between the molecules of liquefied gases. The laws of transition from the gaseous into the liquid state are attributed to Van der Waals and the corresponding binding forces are named after him.

Van der Waals forces are responsible for the bonding at very low temperatures of noble gas atoms—He, Ne, A—into a solid body. Forces of this type also hold together molecules of organic substances to form crystals. These forces are not saturated; the formation of a bond between two atoms does not preclude the binding of a third and fourth atom.

All three types of force are frequently found together in the formation of complex compounds. Figure 54 shows the distribution of electric charge in the compound $CS(NH_2)_2$. The bonds between the atoms C and S and between N and H_2 are of the valence type. The ions $(CS)^{++}$ and $(NH_2)^-$, on the other hand, are held together by ionic forces. The high-density peak in Fig. 54 corresponds to the ion $(CS)^{++}$ and the two lower ones to the two ions $(NH_2)^-$.

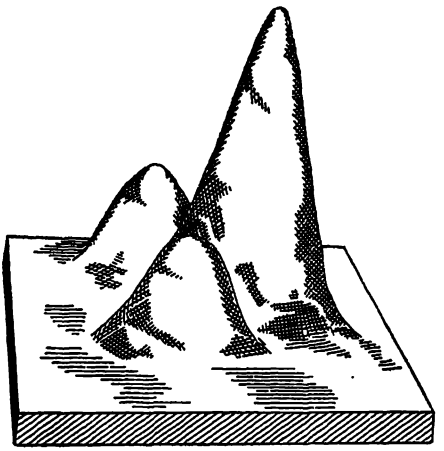


Fig. 54. The distribution of electron density in the compound $CS(NH_2)_2$. The lower peaks correspond to $(NH_2)^-$, the high peak to $(CS)^{2+}$. The bonding within the groups NH_2 and CS is covalent; that between CS and NH_2 is ionic.

The hydrogen bond is due to the attraction towards neighbouring atoms of a hydrogen ion which, in contrast with other ions, consists of a proton deprived of its electron shell. A hydrogen ion may even enter into the electronic shell of a molecule. Hydrogen bonds participate in the formation of water and ice together with the attractive forces between dipoles which exist in H_2O molecules. This type of bond is saturated.

The non-saturated metallic bonds, existing between the totality of positive ions and all the valence electrons, account for the high plasticity and strength of metals. This type of bonding may be found, together with others, in some semiconductors.

From the point of view of their structures crystal lattices are classified according to the presence of elements of symmetry. The highest degree of symmetry is to be found in regular cubic structures, but it should not be inferred that they resemble an isotropic medium.

Figure 55 shows the dependence upon crystallographic orientation of Young's modulus of the crystals of iron, gold and aluminium which

belong to the regular system. Fig. 56 gives the modulus of elasticity for the crystal of magnesium which possesses a uniaxial hexagonal structure.

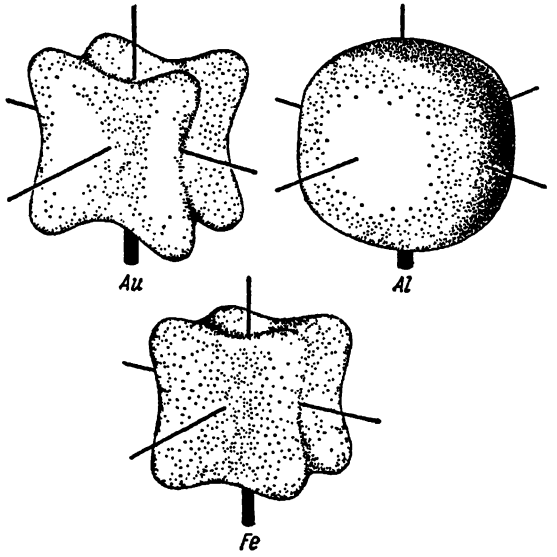


Fig. 55. Diagrammatic representation of the variation of Young's modulus with crystallographic orientation for Au, Al and Fe.

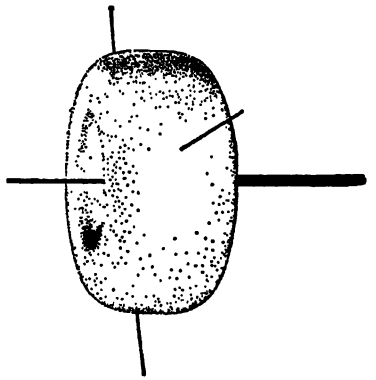


Fig. 56. The modulus of elasticity in Mg crystals for different orientations.

An important characteristic of the properties of the lattice is its coordination number, i.e. the number of nearest neighbours surrounding a given atom. The higher the coordination number the higher the density of the lattice.

Fig. 57(a) shows a lattice with coordination number 4 (ZnS or diamond type); Fig. 57(b) represents the coordination number 6 (simple cubic, NaCl); the coordination number 8 corresponds to a body-centered cubic lattice (Fig. 57(c), CsCl); finally, the coordination number 12 corresponds to a face-centered cubic structure (Fig. 57(d)).

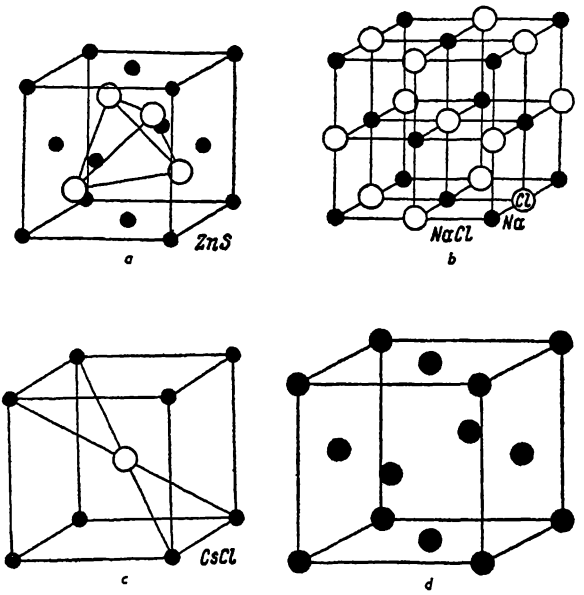


Fig. 57. Crystal lattices with different coordination numbers: (a) ZnS lattice, coordination number 4; (b) NaCl lattice, coordination number 6; (c) CsCl lattice, coordination number 8; (d) face-centred cubic lattice, coordination number 12.

Positive Ions

Li^+	0·60	Be^{++}	0·31	B^{+++}	0·20	C^{++++}	0·15
Na^+	0·95	Mg^{++}	0·65	Al^{+++}	0·50	Si^{++++}	0·41
K^+	1·33	Ca^{++}	0·99	Se^{+++}	0·81	Ti^{++++}	0·68
Cu^+	0·96	Zn^{++}	0·74	Ga^{+++}	0·62	Ge^{++++}	0·53
Rb^+	1·48	Sr^{++}	1·13	Y^{+++}	0·93	Zr^{++++}	0·80
Ag^+	1·26	Cd^{++}	0·97	In^{+++}	0·81	Sn^{++++}	0·71
Cs^+	1·69	Ba^{++}	1·35	La^{+++}	1·15	Ce^{++++}	1·01
Au^+	1·37	Hg^{++}	1·10	Te^{+++}	0·95	Pb^{++++}	0·84

Negative Ions

H^-	2·08						
F^-	1·36	O^{--}	1·40	N^{---}	1·71	C^{----}	2·60
Cl^-	1·81	S^{--}	1·84	P^{---}	2·12	Si^{----}	2·71
Br^-	1·95	Se^{--}	1·98	As^{---}	2·22	Ge^{----}	2·72
I^-	2·16	Te^{--}	2·21	Sb^{---}	2·45	Sn^{----}	2·94

The latter system and the hexagonal structure give the densest packing of atoms and are found in many metals. They are also found

in ionic crystals, which contain ions with completed electronic outer shells and approximately spherical symmetry. Many crystallographic structures may be derived by accommodating the constituent ions in the closest fashion. The solution of this problem requires a knowledge of the dimensions of different ions, some of which are shown in the table above (in Angstrom units).

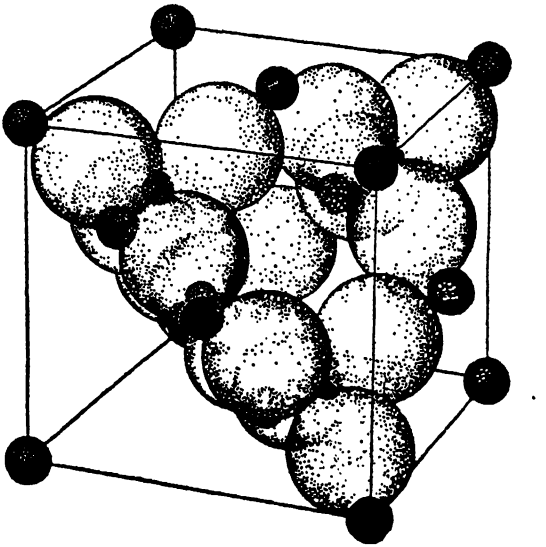


Fig. 58. Close packing of negative ions.

Negative ions are larger than positive ions. For this reason the structure of many oxides and sulphides represents the closest packing of the ions O^{--} or S^{--}, the positive ions fitting into empty spaces between the spherical O^{--} or S^{--} ions. Fig. 58 shows one such compound.

Fig. 59 shows the dimensions of neutral atoms.

A loose structure with the coordination number 4 does, in fact, become denser upon melting and the short-range order gives a higher number of neighbours for each atom. The coordination numbers of Ge and Si change from 4 to 8 on melting. Grey tin, with the coordination number 4, changes to the white form, with the coordination number 6, at temperatures above 13° C. In all these cases the increase of coordination number is accompanied by a transition from semiconducting to metallic properties and a change from the valence to the metallic type of bond. Such a transition also leads to an increase of density upon melting.

The hydrogen-type bond, which is found, for example, in ice crystals, results, like the valence bond, in a relatively loose structure which becomes more dense on melting. Thus, the density of ice is less than that of water, and the densities of solid germanium and silicon also increase on melting.

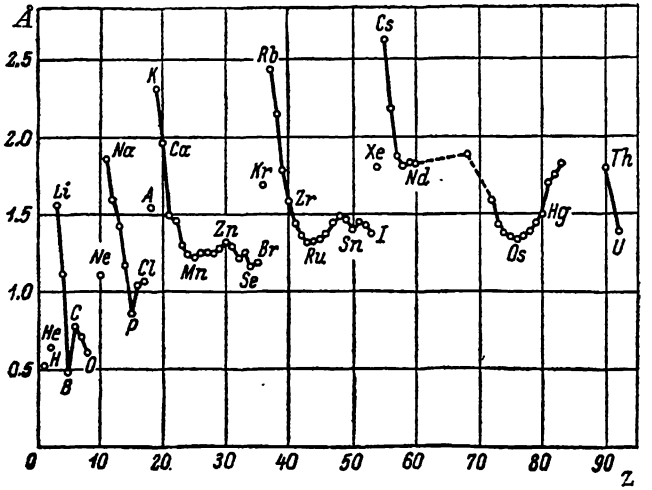


Fig. 59. Atomic radii: Z = atomic number.

Lattices with layer and chain structures show the most pronounced anisotropy. The former cleave readily into thin foils (e.g. mica, graphite, Sb_2S_3, Bi_2S_3, MoS_2). All their properties depend strongly upon the crystallographic orientation. Fig. 60 shows the disposition of carbon atoms in graphite and Fig. 61 the dependence of its thermal conductivity upon orientation; the electrical conductivity varies even more strongly. In the plane parallel to the layers, graphite behaves like a metal, but, in the perpendicular plane, like a semiconductor of high resistivity. Such materials grow in the form of thin single crystal plates when deposited from the gaseous phase or evaporated in vacuo.

Selenium and tellurium crystals are composed of spiral chains or atoms which have strongly covalent bonds, while the chains themselves are bound to each other by weak Van der Waals forces. Modifications of phosphorus, arsenic and antimony form corrugated rather than plane layer structures.

The chemical elements: Be, Mg, Ca, La, Ti, Zr, Hf, Cu, Ag, Au, Zn, Cd, Hg, Ga, In, Tl, Pb, Ru, Rh, Pd, Os, Ir, Pt, Th, U show the closest packing with the coordination number 12; the atoms of a number

of other metals are less densely packed: Li, Na, K, V, Nb, Ta, Cr, Mo, W, α-Fe have a coordination number 8.

The coordination number of valence lattices, such as those of C, Si, Ge, is determined by their valency.

The electronegative elements O, S, F, Cl form molecular structures of, for example, two O atoms, or six or eight S atoms. Fig. 62 shows the structure of a ring of eight sulphur atoms.

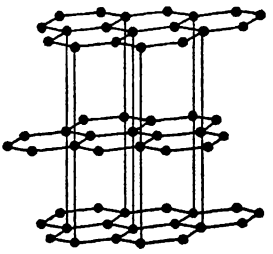


Fig. 60. Graphite lattice.

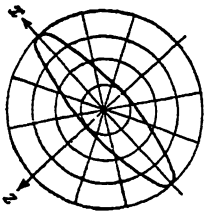


Fig. 61. Anisotropy of thermal conductivity $\lambda_{\perp}$ of graphite.

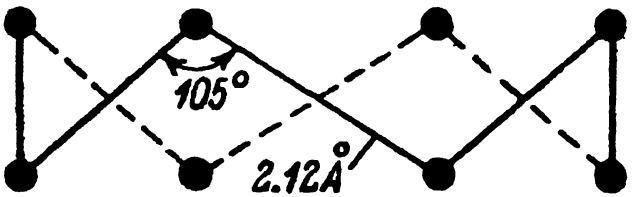


Fig. 62. The molecule S_8.

The noble gases He, Ne, A, Kr in the solid state form regular lattices held together by Van der Waals forces.

The determination of crystal lattice structures is often a complicated task, and is based on a careful analysis of the interference patterns of X-rays, electron waves or neutrons. Each of these methods has its sphere of applications: X-rays are scattered only by electrons, electronic waves only by charged ions, and neutrons by all atoms, irrespective of their charge. The problem is further complicated by the fact that, while it is easy to determine the interference pattern corresponding to a given structure, the converse task of finding the disposition of atoms for their interference maxima can only be accomplished by trial and induction in

the choice of suitable structures. Fig. 63 shows the disposition of atoms in the lattice structures of some semiconductors.

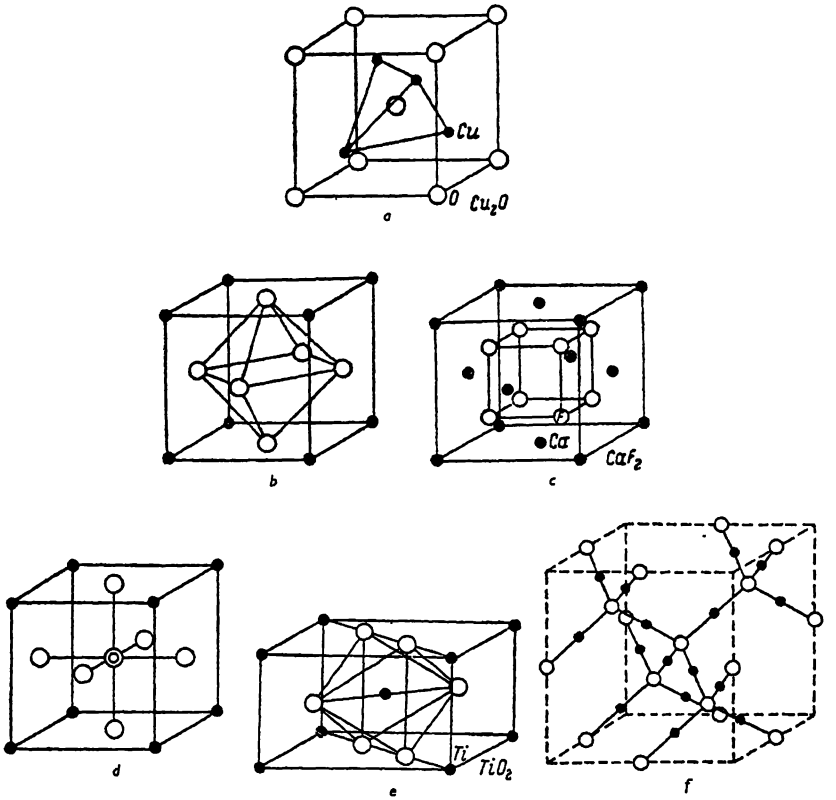


Fig. 63. Crystal lattices of some semiconductors: (a) copper oxide—Cu_2O, coordination numbers 2 and 4; (b) lattice of the type AB_3 (e.g. WO_3); (c) lattice of CaF_2 with coordination numbers 4 and 8; (d) lattice of the type ABC_3 (e.g. perovskite $CaTiO_3$); (e) rutile structure TiO_2, coordination numbers 3 and 6; (f) lattice of cristobalite SiO_2, coordination numbers 2 and 4.

Imperfections in crystal lattices

The strict periodicity of real solids is disturbed by many imperfections or defects. These imperfections have a considerable and, sometimes, even the controlling influence on the mechanical, thermal, electrical and optical properties of semiconductors. They determine the plasticity, hardness, and thermal and electrical conductivities. We shall therefore consider the basic types of imperfections.[2, 3, 4]

Thermal agitation, besides giving rise to statistical deviations of atoms from their equilibrium positions, also causes some permanent changes; occasionally an atom attains so large an amplitude of oscillation

that it leaves its lattice site. If it comes to rest in a neighbouring lattice cell we have, in the crystal, a vacant lattice site and an interstitial atom—a type of imperfection introduced into the theory of solid state by Frenkel'. Schottky assumed that lattice vacancies are formed at the surface and gradually diffuse to the interior of the crystal, while the displaced atoms migrate to the surface. In the case of an ionic crystal, the most probable situation corresponds to the simultaneous formation of two vacant sites due to the migration of one positive and one negative ion.

Both Frenkel' and Schottky defects are found in a number of semiconductors. The density of defects increases with the temperature. At a temperature close to the melting point, the density of such defects in thermal equilibrium may reach 1 to 2% of the total density of atoms in the lattice. As the formation of each defect requires an expenditure of energy to overcome a barrier, an increase in the density of defects leads to an increase in the specific heat. For example, in silver bromide the specific heat increases before melting from 13 to 35 cal/mole.

Another type of disturbance of the periodicity occurs in the formation of solid solutions. Atoms of one substance may distribute themselves among atoms of another substance, either regularly or randomly. The former case corresponds to an ordered alloy structure, the latter to disorder. A critical temperature exists below which the equilibrium structure is in the ordered state and above which the transition to disorder takes place. The difference between these states may be shown by X-ray analysis. The disordered alloy gives the same interference pattern as the basic substance itself. The ordered alloy shows two complementary superstructures of lines corresponding to the two interwoven lattices.

Further imperfections in crystal structures are introduced as a result of the conditions of growth, and dislocations arise due to mechanical stresses. As the crystal grows, dislocations may frequently lead to complicated structures extending over distances of the order of a micron. Such defects may be made visible under an electron microscope and Fig. 64 shows an example of a paraffin crystal near a dislocation. Two classes may be distinguished: edge dislocations and screw dislocations. In crystal growth the formation of a new facet requires some additional energy, in excess of that needed for the continuation of the existing facet. Spiral growth has the advantage that it proceeds by continuously extending the same atomic layer.

The presence of dislocations drastically alters the mechanical properties of a crystal, increasing its strength by a factor of a hundred or so. Although dislocations do not markedly affect the mobility of electrons, they reduce the electron diffusion length.

Semiconductors are frequently prepared and investigated, not in single crystal form, but as polycrystalline aggregates, compressed powders or sintered masses. Sometimes, also, porous or non-uniform specimens consisting of two or more mixed phases are prepared.

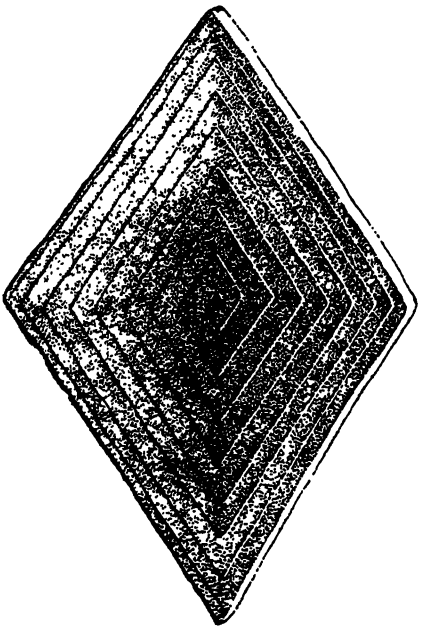


Fig. 64. Growth spiral in a paraffin crystal.

The properties of such specimens obviously differ from the properties of a single crystal specimen of the same material. Each grain boundary in a polycrystalline material disturbs the periodicity of the lattice. If the grains are anisotropic and are differently orientated, the boundaries give rise to the scattering of acoustical vibrations and a reduction of the thermal conductivity. Impurities diffuse, and gases are transported along grain boundaries. Oxygen may be adsorbed on the surface of grains in the form of negative ions or as an oxide layer. Oxide barriers having a high resistivity appear in the solid and reduce the electrical conductivity of the specimen, notwithstanding their small thickness. Such barriers have only a very slight effect on the thermal conductivity of the specimen, since their own specific thermal conductivity may differ from that of the basic material only by one or two orders of magnitude. If the thickness of the barriers is of the order of 10^{-5} to 10^{-6} cm and the dimensions of crystallites are 10^{-2} cm, then the effect of the barriers on the thermal conductivity is negligible. In the case of electrical conduction, however, the ratio of the resistivities of the barrier and the basic material may

reach 10^{10} and, thus, with the above dimensions the resistance may increase a million times.

In such cases it is particularly important to have single crystal material both for applications and in research.

Some materials are easily obtainable in the form of large crystals by growth from solutions. Examples of such are sulphur, Rochelle salt, alums and some salts.

Plate- and needle-shaped crystals are obtainable by evaporation *in vacuo*.

The main method of obtaining semiconducting materials in a single crystal form consists, however, in growing from the melt. Crystals of germanium, silicon and of a number of compounds are obtained in this way. Several methods have been proposed. Czochralski pulled the crystal seed from the melt at such a speed that the interface of the growing crystal was always level with the liquid. Obreimov placed the melt in a tube with one end drawn out and gradually cooled it from that end, thus initiating the growth of crystals. Of these crystals, the one with the orientation corresponding to the direction of most rapid growth soon overtakes the others and continues to grow as a single crystal.

Sometimes, on cooling slowly, instead of a single crystal, one obtains a system of crystallites similarly orientated in the direction of most rapid growth—this is the case of orientated crystal aggregates.

Impurities

Impurities have a decisive influence upon the electrical properties of semiconductors: they alter the resistivity by a factor of a thousand or even a million times. Impurities determine the magnitude of the conductivity and the sign of current carriers; the mean free paths of electrons and phonons depend largely upon the density and type of impurities which, thus, determine the mobility, the thermal conductivity and the diffusion length in semiconductors.

The term impurity is taken to comprise not only actual admixtures of foreign matter in a crystal lattice but is given a broader meaning, including departures from the stoichiometric proportions in the composition of compound lattices. Thus, an excess of oxygen ions or a deficit of copper ions in copper oxide may be regarded as an admixture of oxygen or as an admixture of a certain number of divalent copper ions taking the place of the monovalent ions or, finally, as the presence in the lattice of empty sites vacated by the migrating copper ions. All these will be covered by the term impurities. They cause the conductivity of copper oxide to be of the order of 10^{-3} to 10^{-4} ohm^{-1} cm^{-1}, while stoichiometric copper oxide with no impurities, would have a conductivity of the order 10^{-9} ohm^{-1} cm^{-1}.

From this point of view one should distinguish between the following impurities: (1) those representing an excess or deficit of one of the components of the basic lattice; (2) the substitution of another atom or ion in a lattice site belonging properly to a constituent of the lattice (this type is called a substitutional impurity); (3) foreign atoms present interstitially within lattice cells (these are termed interstitial impurities).

Departure from the stoichiometric composition gives rise to the solution of a phase richer in the excess component, in another phase having a deficit of the component in question. The solubility of one phase in the other, i.e. the maximum concentration of the solution, is usually limited to a small range of the order of 1%. Nevertheless, if the electrons of the impurity atom have a low energy of dissociation it may have a decisive influence. Thus, with a forbidden gap width of 1 eV, the intrinsic carrier concentration at room temperature amounts to 10^{12} cm^{-3}, while 0·1% of an impurity, with an activation energy of 0·25 eV, gives a carrier concentration of 10^{17} cm^{-3}.

The activation energy of impurity atoms may be determined in the following manner. Let the ionization energy of the atom *in vacuo* be U. In a medium of dielectric constant ϵ, the dimensions of the electronic orbits increase by a factor ϵ. For large values of ϵ these dimensions are considerably greater than the inter-atomic distance a. Thus, in silicon with ϵ equal to 13·5, the radius of a univalent orbit is about 15 Å, i.e. 3 to 5 times the distance a. This allows one to consider the medium surrounding the electron as continuous and the electric field as very nearly the Coulomb field, reduced by a factor ϵ. Thus the ionization energy is reduced by ϵ^2 and becomes U/ϵ^2. In the case of silicon, ϵ^2 is 170. If the impurity atom has a hydrogen-like electronic orbit with an ionization energy of 13·5 eV in free space, the ionization energy in silicon would amount to 0·07 eV. Experimental results fully confirm the validity of this rather crude model, which interprets the transition into the conduction band as being equivalent to ionization.

The sign of the current carriers due to an impurity depend on its chemical nature, its valency and on the manner of its incorporation in the lattice.

In the case of interstitial impurities one may anticipate that electropositive atoms which readily give off their electrons, with an ionization energy U/ϵ^2, will give rise to energy levels close to the bottom of the conduction band (assuming reasonable values for ϵ), and the resulting conductivity will be due to electrons, i.e. n-type. Electro-negative atoms which require a moderate energy U' to assimilate an extra electron will give rise to levels at a small distance U'/ϵ^2 above the top of the valence band and will result in p-type conductivity. In this manner electropositive interstitial impurities become donors and electro-negative ones become acceptors.

Different conditions obtain for substitutional impurities in valence crystals. Every atom entering the lattice contributes one electron to each of the valence bonds with neighbouring atoms. The coordination number of the lattice is usually determined by the valency of its constituent atoms. Thus, the atoms of tetravalent carbon (diamond), silicon, germanium and grey tin each have four nearest neighbours and use four valence electrons in their bonds. If one of these atoms is replaced by another of different valency, then its bonding in the lattice still requires four, and only four, electrons.

When the impurity atom has a higher valency, as is the case with the pentavalent elements arsenic, phosphorus and antimony, its excess electron is not used in the formation of bonds. When the medium has a high dielectric constant, this atom has a low activation energy U/ϵ^2 and it represents a donor.

Atoms with a lower valency than the host atoms, for example trivalent aluminium, gallium and indium, cannot make good the deficit from their own inner electronic shells as this requires a very high energy (the dissociation energy of inner-shell electrons considerably exceeds the ionization energy of the valence electrons). A vacant electronic site (hole) is left instead in the valence band.

In this way substitutional impurities in valence lattices act as donors if the number of their valence electrons is higher than that of the lattice, and as acceptors if it is lower. In contrast to the case of interstitial impurities, electro-positive impurities are acceptors and electro-negative impurities are donors.

We expect that in ionic lattices the sign of the current carriers is determined by the chemical nature of the impurities. A neutral atom may either give off an electron into the conduction band and become a positive ion, or it may take up an electron from the valence band and become a negative ion in the process.

These simple considerations cease to apply in valence semiconductors with a low dielectric constant. The periodic field of the lattice cannot then be replaced by a continuous medium and the modification of the filled levels caused by the introduction of the impurity must be considered in greater detail.

Some information about the nature of the disturbance introduced into the electric field by various structural defects may be obtained from measurement of their influence on the high frequency paramagnetic resonance spectrum.

The magnitude of these distortions may also be inferred from their effect on the mobility of electrons and on the mean free path of phonons. Substitutional impurities distort the lattice but their influence is confined to the nearest lattice cell. Interstitial atoms distort the lattice over

distances extending to several interatomic spacings, if the impurity atom has a large radius.

In view of the considerable influence of impurities, the chemical purification of semiconductors assumes a major importance. This problem presents requirements exceeding anything known before in analytical chemistry. Hitherto a chemical reagent was regarded as pure if its impurity content was below 0·01 to 0·001%. However, in order to achieve "purity" in a material such as silicon, with an energy gap ΔE_0 equal to 1·12 eV, or germanium, with ΔE_0 equal to 0·75 eV, it is necessary to reduce the impurity content to 10^{-10} to 10^{-7}%. Only then is the concentration of impurity electrons or holes small compared with the intrinsic carrier concentration in the material.

In fact, the intrinsic concentration n_0 of electrons and holes at 300° K is about 10^{10} cm^{-3} for ΔE_0 equal to 1·12 eV and about 10^{13} cm^{-3} for ΔE_0 equal to 0·75 eV. Thus, the concentration N of impurities should be no higher than these figures so that the impurity carriers should not be more numerous than the intrinsic carriers. In this consideration we assume that, since the dissociation energy of impurities in these substances is much less than kT, all impurity atoms are fully ionized. The total density of atoms N_0 is about 10^{22} cm^{-3}, so the requirement is that N/N_0 should amount to 10^{-10}% in the first case, and 10^{-7}% in the second case. Even at these low concentrations the impurity carriers are still equal in density to the intrinsic carrier densities. A still higher purity would be required if the impurity concentration should remain negligible.

New technological methods were required to achieve these orders of purity. One of these consists in repeated crystallization from the melt. The solubility of impurities, if present in atomic form, is usually less in the solid crystal than in the melt with the result that, on each recrystallization, the crystal appears purer than the melt. Successive melting and renewed pulling of single crystals from the melt enables the gradual reduction of the impurity content to be achieved. This operation is simplified in the method of zone refining: a relatively narrow molten zone is formed in a long ingot by means of a high frequency or radiation furnace. As the hot zone is moved along the ingot successive parts become molten and then freeze again until the end of the ingot is reached. The concentration of impurities is increased in the direction of zone movement. By a repeated application of this process it is possible to obtain a crystal in which most of the impurity is concentrated at one end and the purity increases with distance towards the other end.[5]

The purity requirements imposed on the material are so stringent that it is important to guard against contamination by diffusion of impurities from the crucible containing the melt or the boat in which the zone melting is carried out. Atmospheric pollution in the laboratory

may introduce detectable levels of impurity. Thus, e.g. oxygen and copper readily enter into, and rapidly diffuse through, germanium and are always detectable as impurities.

Impurities do not only represent a disadvantage but may give the semiconductor highly desirable properties. In order to achieve this end it is first necessary to purify the semiconductor to the highest practicable extent and, thereafter, to introduce a controlled concentration of the desired impurities. In this way it is possible to introduce a p-type layer into an n-type semiconductor or even to obtain a series of p- and n-type layers.[6]

Of particular importance are semiconductors with compensated donor and acceptor impurities. The donor levels are situated near the edge of the conduction band. They lie above the acceptor levels, situated near the valence band. Therefore, in thermal equilibrium, electrons from the donor levels fall into the empty acceptor levels. If a semiconductor contains N_1 donors and N_2 acceptors per cc, then the net concentration is (N_1-N_2) electrons in the donor levels if $N_1>N_2$, or (N_2-N_1) empty acceptor levels if $N_2>N_1$. If N_1 equals N_2 then there are neither electrons in the donor levels nor empty acceptor levels. From the point of view of the current carrier concentrations, such a semiconductor shows the same behaviour as if it were perfectly pure.

The concentration of holes or electrons in this case is:

$$n_0 = \frac{2(2\pi m^* kT)^{\frac{3}{2}}}{h^3} \exp\left(-\frac{\Delta E_0}{2kT}\right) = N_{\text{eff}} \exp\left(-\frac{\Delta E_0}{2kT}\right) \qquad (95)$$

as for the complete absence of impurities.

The presence of a large concentration of compensated impurities does, however, affect other properties: the mobilities of the current carriers are reduced as compared with the pure material. The mechanism of scattering is altered; in the pure substance the scattering of electrons by thermal vibrations increases with increasing temperature, while the scattering by positive donor ions and negative acceptor ions, which may predominate in a semiconductor with compensated impurities, decreases with rising temperature. In the absence of degeneracy, the velocity of electrons is proportional to the square root of the temperature and the angle of deflection suffered by electrons in passing through the Coulomb fields of ions is accordingly reduced at high temperatures.

The presence of impurities also reduces the diffusion velocity and diffusion lengths of electrons and holes, and reduces the mean free path of phonons and, thus, the thermal conductivity of the lattice.

It is relatively easy to add both donor and acceptor impurities to the following semiconductors: Si, Ge, PbS, PbSe, PbTe, SiC, Tl_2S, U_3O_4 and UO_2.

Semiconductors with predominantly *n*-type conductivity are: TiO_2, V_2O_5, CdS, CdSe, HgS, Hg_2S, CdO, ZnO, Ag_2S, CsS, WO_3 and Al_2O_3, while Cu_2O, Ag_2O, Cu_2S, Sb_2S_3, Te, Se, Hg_2O, MnO, CoO, NiO and SnO are *p*-type.

Thermal vibrations of the lattice

The thermal motion of the lattice has a considerable influence upon the motion of electronic charges, i.e. on their mobility. Moreover, both the energy level structure and, in particular, the width of the forbidden gap vary with temperature, since the temperature affects the interatomic distance which, in turn, determines the energy levels. For the majority of semiconductors the width of the forbidden gap changes by some $(2\text{–}4)\times 10^{-4}$ eV/° C. Typical values are: $-(3\text{–}5)\times 10^{-4}$ for Si, $-(1\text{–}5)\times 10^{-4}$ for Ge, and $+2\times 10^{-4}$ for Te. In cuprous oxide, cooling through 200° C, from room temperature to liquid air temperature, shifts the absorption spectrum towards shorter wavelengths and increases the gap width by 0·2 eV. We propose to discuss the problem of thermal vibrations in a little more detail than has been done so far.

Mathematical analysis shows that the random thermal vibrations of atoms in a crystal lattice may be represented by a sum of all the possible normal modes, or standing waves, in the lattice. A condition for the existence of a standing wave is that an integral number of half-waves should be present in its path from one boundary of the solid to the other. We assume, for the purpose of the analysis, that the solid is in the form of a rectangular parallelepiped with edges a, b, c. It is evident, however, that the thermal vibrations cannot depend upon the geometrical shape and dimensions of the solid. We shall consider one of the standing plane waves making angles $\alpha_1, \alpha_2, \alpha_3$ with the edges of the body. The condition that the path of the wave should contain an integral number of half-waves may be expressed by the equations:

$$a\cos\alpha_1 = m_1\frac{\lambda}{2}$$

$$b\cos\alpha_2 = m_2\frac{\lambda}{2}$$

$$c\cos\alpha_3 = m_3\frac{\lambda}{2} \tag{96}$$

where m_1, m_2, m_3 are integers.

However, for any arbitrary direction,

$$\cos^2\alpha_1+\cos^2\alpha_2+\cos^2\alpha_3 = 1,$$

and hence:

$$\frac{m_1^2}{4a^2}+\frac{m_2^2}{4b^2}+\frac{m_3^2}{4c^2} = \frac{1}{\lambda^2} \tag{97}$$

$m_1/2a$, $m_2/2b$ and $m_3/2c$ may be regarded as three components of the vector $1/\lambda$. By giving all possible values to m_1, m_2 and m_3 we obtain all the normal vibrational modes of the solid.

Let all values of $m_1/2a, m_2/2b, m_3/2c$ be plotted in a three-dimensional rectangular coordinate system (Fig. 65). The number of points with such coordinates in a volume $(m_1 m_2 m_3)/(8abc)$ gives the number of values of $1/\lambda$.

The distance between two neighbouring points along the ξ-axis, corresponding to values of m_1 differing by unity, is equal to $1/2a$, along the η-axis the distance is $1/2b$, and along the ζ-axis is $1/2c$. If the interatomic distance is $2d$, then the standing wave length, equal to $\lambda/2$, must be greater than d and $1/\lambda$ must be less than $1/2d$.

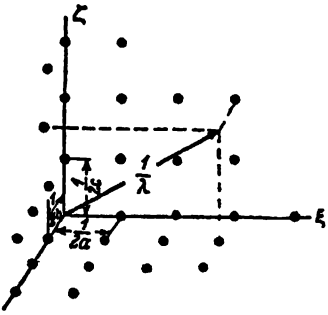


Fig. 65. The possible modes of vibration of a rectangular solid.

By this process the task of determining the total number of standing waves in the solid is reduced to the computation of the total number of points, denoting the various possible values of $1/\lambda$, for which $\lambda > 2d$ or $1/\lambda < 1/2d$. The total number of possible waves will be given by the number of points enclosed by a sphere of radius $1/2d$ in our coordinate system. The volume of this sphere is $4\pi/3(2d)^3$, that is $\pi/6d^3$.

For an arbitrary wavelength λ_0, the number of normal modes for which $\lambda > \lambda_0$, or $1/\lambda < 1/\lambda_0$, is

$$n = \frac{\pi}{6\lambda_0^3} \tag{98}$$

and the number of modes of vibration lying between λ_0 and $(\lambda_0 + d\lambda)$ is given by:

$$\frac{dn}{d\lambda} = \frac{\pi}{2\lambda_0^4}. \tag{98a}$$

Going over from the wavelength λ to the frequency of vibrations, using the relation

$$h\nu = v,$$

where v is the velocity of propagation of waves in the solid, i.e. the velocity of sound, we obtain

$$n = \frac{\pi\nu^3}{6v^3}, \quad \frac{dn}{d\nu} = \frac{\pi\nu^2}{2v^3}. \tag{99}$$

In this derivation, as well as in what follows, we shall make the simplifying assumption that the velocity of sound v does not depend upon frequency, i.e. we shall neglect the dispersion of sound waves, down to wavelengths of atomic dimensions $2d$.

For a wave travelling at an angle α to the plane of the lattice, the minimum value for the wavelength is $2d\cos\alpha$ and the maximum value of $1/\lambda$ is $1/(2d\cos\alpha)$. Plotting this value of $1/\lambda$ for various angles α, we may define the volume containing all the possible values of $1/\lambda$ for a given lattice. The figure so obtained is the same Brillouin zone which we have introduced in the quantum theory of electronic energy levels, in the coordinate system $1/2a$, $1/2b$, $1/2c$. It is also evident that, as long as we are considering only the wavelengths, the geometrical conditions for the reflection of electronic and acoustic waves are identical. However, the difference between the velocity of sound (of the order of 10^5 cm/sec) and the velocity of propagation of electronic waves (of the order 10^8 cm/sec) leads to major differences between the frequencies ν corresponding to a given wavelength λ.

The geometrical conditions for reflection of acoustic and electronic waves are the same as the laws governing the reflection of X-rays. In order to determine the shape of the Brillouin zone, it is possible, therefore, to use the same method of reciprocal lattices, with coordinates $m_1/a, m_2/b, m_3/c$, as has been employed in the determination of X-ray patterns.

The method of reciprocal lattices not only leads to a value for the total number of different vibrations, which should be $3N$ where N is the number of atoms, but also gives the distribution of vibrations according to wavelength and, consequently, also according to frequency. To each wavelength in a monatomic lattice corresponds one longitudinal wave and two transverse waves polarized in perpendicular planes.

The probability W that a mode of oscillation with frequency ν should be excited at a temperature T is the same as the probability of finding a photon of frequency ν in the spectrum of radiant energy and is given by:

$$W = \frac{1}{\exp(h\nu/kT) - 1}.$$

The total energy of the thermal vibrations contained in an interval $d\nu$, between the frequencies ν and $(\nu + d\nu)$, is obtained by multiplying

the number dn of modes in that interval by the probability W of finding these modes excited and by the energy $h\nu$ of the sound waves (phonons).

Born and von Karman in 1912 gave a strict mathematical derivation of the frequency distribution of energy for a cubic lattice. The spectrum of vibrations which they obtained was rather complicated, consisting of five branches, each of them cut off at its limiting frequency ν_i.[7]

As early as 1906 Einstein introduced quantum-mechanical considerations into the theory of thermal vibrations, starting from the simplified hypothesis that all atoms in the solid vibrate at one frequency ν_0. In 1911 he proved that the thermal energy of a solid containing N atoms at a temperature T is

$$3Nh\nu_0/[\exp(h\nu_0/kT)-1].$$

It is necessary to introduce into this formula a correction corresponding to the energy $\frac{1}{2}h\nu_0$ which remains in each vibration at the absolute zero:

$$3N\left[\frac{h\nu_0}{\exp(h\nu_0/kT)-1}+\frac{h\nu_0}{2}\right]. \tag{100}$$

In 1913 Debye computed the thermal energy Q_T and specific heat c_v of a solid by replacing the crystal lattice with a continuous medium possessing the same velocity for sound waves as the crystal. Debye took account of the atomic nature of the solid by limiting the total number of vibrations to $3N$.

The only constant which determines the specific heat of a solid according to Debye's theory is the maximum frequency ν_{max}, corresponding to the minimum wavelength which is equal to $2d$ in an atomic lattice. In Debye's theory ν_{max} is related to the velocity of sound. Even in isotropic media the velocity $v_\parallel$ of longitudinal waves differs from the velocity $v_\perp$ of transverse waves. Debye assumed one mean value for the velocity v given by the relation:

$$\frac{3}{v^3}=\frac{1}{v_\parallel^3}+\frac{2}{v_\perp^3}. \tag{101}$$

These considerations led to the result that the temperature dependence of the specific heat differs only in the scale of temperature T. It is possible to define a temperature θ for every solid such that the dependence of the specific heat upon (T/θ) is the same for all solids. If T is measured in units of θ, i.e. if T/θ is chosen as the independent variable, the thermal energy Q_T per gram-atom of the solid is given by the relation:

$$Q_T = 3RTD(T/\theta)$$

where R is the gas constant per gram-atom, equal to 1·98 cal/degree, and

$$D\left(\frac{T}{\theta}\right)=3\left(\frac{T}{\theta}\right)^3\int_0^{\theta/T}\frac{x}{\exp(x)-1}x^2\,dx \tag{102}$$

where

$$x = \frac{h\nu}{kT}.$$

The temperature θ is called the Debye temperature and is determined by the maximum frequency ν_{max}:

$$\theta = h\nu_{max}/k.$$

The specific heat per gram-atom is equal to the derivative of energy with respect to temperature and is given by:

$$c_v = \frac{dQ_T}{dT} = 3R\left[4D - \frac{3x}{\exp(x)-1}\right]. \tag{103}$$

For small values of x, i.e. high values of T/θ, $x/[\exp(x)-1]$ is equal to unity, D equals unity, and Q_T equals $3RT$, that is about $6T$ cal. The value of the specific heat c_v is about 6 cal/degree and is in good agreement with the empirical relation obtained by Dulong and Petit. At high values of x, i.e. low temperatures T/θ, we have:

$$Q_T \propto T^4, \quad Q_T = \frac{3\pi^4}{5} R\frac{T^4}{\theta^3}, \tag{104}$$

$$c_v \propto T^3, \quad c_v = \frac{12\pi^4}{5} R\frac{T^3}{\theta^3}. \tag{105}$$

Fig. 66 shows the temperature dependence of the specific heat c_v according to the theories of Debye and Einstein.

The following table gives the Debye temperatures in degrees Kelvin for a range of materials.

A	85° K	Fe	420° K	Pb	88° K
Ag	215	Ga	290	Pt	225
				Sn	260
Al	398	Ir	283	Ta	245
Au	180	K	99	W	310
Be	1000	Li	380–430	Zn	235
C	1860	Mg	290	NaCl	281
Ca	230	Mo	379	KCl	230
Cd	160	Na	159	KBr	177
Cr	485	Ne	63	CaF_2	474
Cu	315	Ni	370	FeS_2	630

Debye's theory analyses the propagation of elastic waves in a monatomic solid. It shows satisfactory agreement with experimental data for a number of monatomic metals and some inorganic compounds, as shown in Fig. 67.

Frequently, however, the results of measurements of the specific heat over a wide range of temperatures cannot be expressed by Debye's formula. This is apparent from the fact that different values of θ may

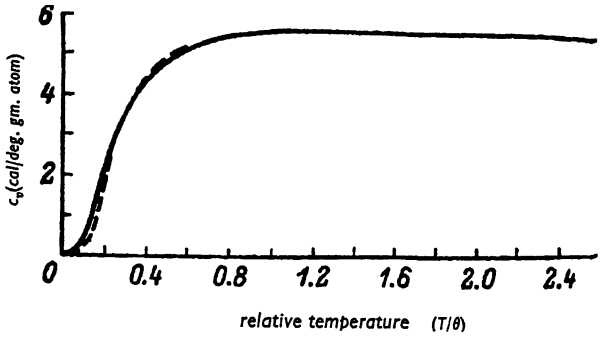


Fig. 66. Comparison of Debye's (solid line) and Einstein's (dotted line) theoretical dependence of the specific heat on temperature.

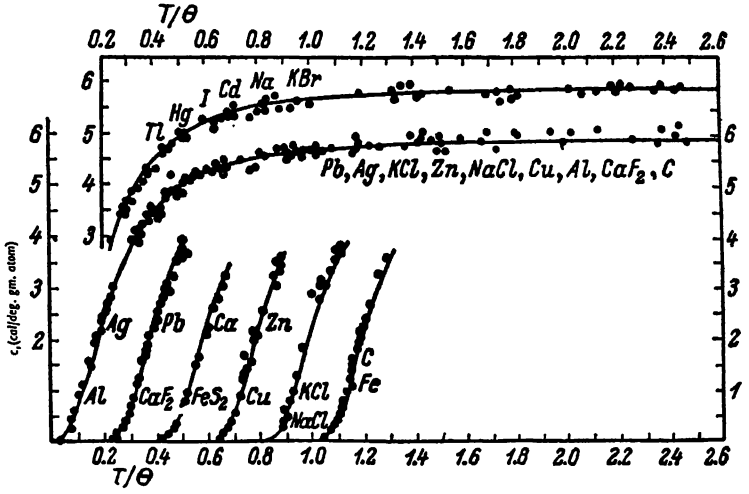


Fig. 67. Experimental results and Debye's theoretical curves for the specific heat.

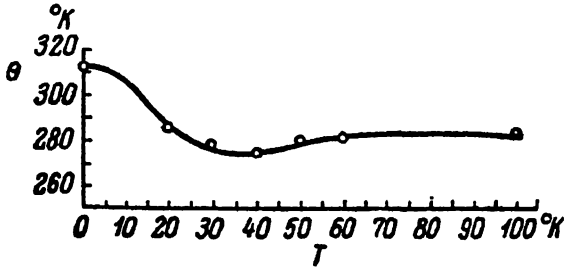


Fig. 68. Example of variation of Debye temperature with temperature.

be obtained for different temperature intervals. In order to bring the experimental results into line with theory it is necessary to regard θ itself not as a characteristic constant of the material but as a function of T (Fig. 68). The following table gives the values of θ for some substances at $T = \theta$ and $T = \theta/12$.

	$T = \theta$	$T = \theta/12$
Au	180	162
Cu	315	319
Li	430	356
Na	150	160
W	310	337

The spectrum of the thermal vibrations of a real solid (i.e. the number of vibration modes contained in the interval ν, $(\nu + d\nu)$) differs considerably from the parabolic relation given by equation (99) and shown in Fig. 69. The actual spectrum of elastic vibrations for simple cubic lattice is shown in Fig. 70, for a body-centred cubic lattice in Fig. 71, and for a face-centred lattice with close atomic packing in Fig. 72.

Both Einstein and Debye took into consideration only the elastic vibrations of the atoms. In molecular lattices, vibrations of atoms within the molecules appear, as well as elastic vibrations of the molecules. At high temperatures the electrons begin to take part in the thermal motion. In the case of ionic crystals, those vibrations in which oppositely charged ions are displaced with respect to each other give rise to electromagnetic waves affecting the mobility of electrons to a marked extent.

It is necessary, therefore, in the first place to distinguish two kinds of vibrations. The first kind, which give rise to electromagnetic waves are called the optical vibrations. The second kind, called acoustic vibrations, are due to displacements of electrically neutral atoms, or to the movement of whole groups of ions with compensating charges which merely changes the local density of the medium. Intramolecular vibrations and vibrations in complex lattices, made up of different types of atoms, are also called optical vibrations in view of their properties.

Among the optical vibrations, which are most pronounced in ionic solids (i.e. in polar ionic crystals, Fig. 73), one should distinguish transverse and longitudinal vibrations. In the case of transverse vibrations, different charges moving in opposite directions give rise to periodically variable electric fields in the space between the neighbouring charges. Since the electronic wavelength corresponding to an energy 3×10^{-2} eV is 7×10^{-7} cm, i.e. it exceeds the inter-atomic distance by more than an order of magnitude, the transverse vibrations affect the electronic motion only slightly. In the case of longitudinal vibrations, however, a group of positive ions is displaced in one direction while a group of negative ions moves in the opposite sense; this results in the

formation of localized electric fields extending over several inter-atomic distances. Random electric fields extending over distances exceeding λ have a very strong deflecting and scattering influence on the electrons.

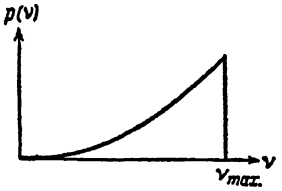


Fig. 69. Spectrum of normal modes of vibration in an atomic lattice according to Debye; $p(\nu) \propto \nu^2$.

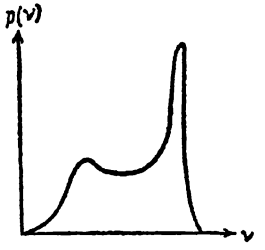


Fig. 70. Vibration spectrum of a cubic lattice.

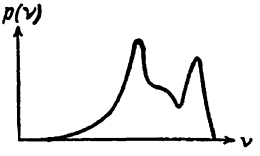


Fig. 71. Vibration spectrum of a body-centred cubic lattice (tungsten).

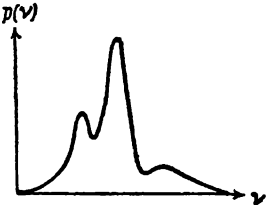


Fig. 72. Vibration spectrum of a face-centred cubic lattice (sodium chloride).

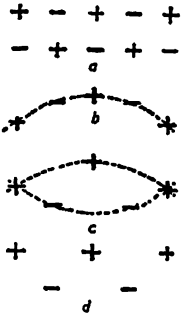


Fig. 73. Different types of vibration in a heteropolar crystal of the rock salt type: (a) distribution of ions in the lattice; (b) acoustic vibration; (c, d) optical vibrations.

Optical vibrations may arise even in crystals made up of uncharged atoms if several atoms enter into one unit cell.

Only in crystals belonging to the regular system are the frequency and the velocity of propagation of the acoustic vibrations the same in all directions. In other crystals the velocity of sound may be represented by an ellipsoid. The velocity of sound may be considered to be practically independent of the frequency of vibrations down to wavelengths of the order of inter-atomic dimensions. Each atom gives rise to one longitudinal vibration with an appropriate velocity of propagation and two mutually perpendicular transverse vibrations propagated with a different velocity. The total number of possible vibrations in a solid consisting of N atoms is $3N$.

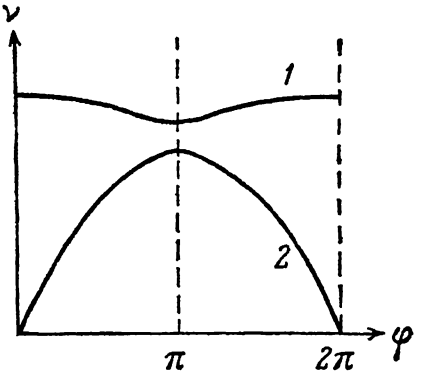


Fig. 74. Relation between frequency ν and the phase difference ϕ for (1) optical vibrations and (2) acoustic vibrations.

All possible vibrations in a solid may be distinguished according to the magnitude of the phase difference ϕ between neighbouring atoms. The magnitude of ϕ is the higher, the shorter the wavelength. The very shortest wavelengths have a phase difference ϕ_0 equal to π.

In the case of acoustic vibrations the frequency of vibrations ν increases with increasing ϕ up to a value ν_{max} when ϕ_0 equals π.

Transverse optical vibrations may arise in a lattice consisting of ions or dipolar molecules. The frequency of the vibrations changes only very little with increasing magnitude of the phase difference ϕ between the two neighbouring charges; it decreases slightly at first, reaching a minimum at ϕ equal to π. The minimum of the frequency of the optical vibrations is the more strongly pronounced the higher the mass difference between the positive and negative ions. In the case of equality of the ionic masses, the minimum frequency of optical vibrations at ϕ equal to π coincides with the maximum value for acoustic vibrations at the same value of ϕ.

Fig. 74 shows the relation between ν and ϕ for optical and acoustic vibrations.

Transverse optical vibrations may be excited in a crystal lattice not only by means of thermal agitation but also by infra-red electromagnetic waves. The wavelength of infra-red radiation, corresponding to the maximum frequency ν of ionic vibrations equal to 10^{13} sec^{-1}, amounts to tens of microns, that is some 10^5 times the inter-atomic distance, which is a few Angstrom units. The phase difference for optical vibrations excited by such radiation is therefore very small. An ionic crystal lattice absorbs in that range of frequencies which corresponds to small values of ϕ or relatively long wavelengths. Infra-red radiation of these wavelengths is not only rapidly absorbed, but also particularly strongly reflected. Thus, after only a few reflections from the surface of a heteropolar crystal, the spectrum of the resulting radiation contains only the frequencies corresponding to those wavelengths. These are called the "residual rays" or reststrahlen of the given crystal.

In crystals belonging to the cubic system, the residual rays have the same frequency whatever the crystallographic orientation. However, in crystals with lower symmetry—uniaxial and biaxial—there are two or three frequencies for the residual rays.

The frequency of longitudinal vibrations is higher than that of transverse vibrations:

$$\nu_{\text{long}}^2 = \nu_{\text{trans}}^2 + \frac{e^2}{2\pi M d^3} \tag{106}$$

where M is the reduced ionic mass and d the distance between positive and negative ions.

The totality of all the normal modes of vibration of $2N$ atoms in a polar lattice may be considered as a function of the phase difference between neighbouring ions or as a function of the reciprocal wavelength of the standing waves (which amounts to the same thing). All normal modes fall into separate branches: (1) $3N$ acoustical vibrations of which $2N$ are transverse and N longitudinal; (2) $2N$ transverse and N longitudinal optical vibrations. Each of these types of vibrations gives rise to fluctuations of density or an electrostatic field in the crystal causing a certain amount of scattering of electrons. The same scattering process also maintains equilibrium by removing any departures from normality of the velocity distribution function of the electrons. The highest electric fields occur in ionic crystals and the most pronounced changes of density over considerable volumes are caused by longitudinal vibrations. Thus, the latter, in ionic crystals give rise to a very low electron mobility.

The thermal motion increases with temperature and all types of vibration are intensified. Below the Debye temperature, only the lower frequency vibrations participate in the thermal motion. In this range a change of temperature alters not only the intensities, but also the

frequency spectrum, of the acoustic vibrations of the lattice, in accordance with the laws derived by Born, von Karman and Debye. The optical vibrations have frequencies confined to narrow intervals and are better described by Einstein's theory of specific heats, according to which all vibrations have the same frequency ν_0.

Thus, in lattices in which thermal motion consists of acoustical as well as optical modes, the frequency spectrum may be represented approximately as the combination of a Debye spectrum (Fig. 69) with $3N$ additional vibrations of one frequency ν_0 corresponding to the mean frequency of the optical branch. The temperature dependence of the specific heat in this case is intermediate between the two curves in Fig. 66.

The splitting of the vibration spectrum into two branches, analogous to the case of Fig. 74, also occurs for lattices made up of neutral atoms of two unequal masses M and m and, again, in the case where several atoms enter into one unit cell. In these cases, too, the terms acoustic and optical are retained to denote the two branches. The frequency ν_0 of optical vibrations is expressed, in this case, by the relation:

$$\nu_0 = \frac{1}{2\pi}\left[2\beta\left(\frac{1}{m}+\frac{1}{M}\right)\right]^{\frac{1}{2}} \tag{107}$$

where β is the modulus of elasticity.

If, instead of considering a three-dimensional lattice, we were to deal with a linear chain of atoms bound by elastic forces, the expression for specific heat c_v in the neighbourhood of absolute zero would be proportional to the absolute temperature T rather than the third power of T. Fig. 75 shows the temperature dependence of c_v, with θ equal to 200° K, for a linear chain of atoms, and the difference between the specific heats for periodic and continuous media.

In two-dimensional lattices, at low temperatures, c_v is proportional to $(T/\theta)^2$.

The difference between the dependence of c_v upon T/θ in one- and two-dimensional, as against three-dimensional, lattices, persists even at very high temperatures.

An approximation to a one-dimensional lattice is represented by those lattices in which thread-like atomic chains connected internally by strong valence bonds are joined one to another by relatively weak Van der Waals forces. Linear polymers also form thread-like crystals.

Lattices with layer structures such as graphite and mica are analogous to the two-dimensional case.

It should be understood that the analogy in these cases is only partial since the difference between the vibrations in the three directions is only quantitative. At temperatures near the absolute zero, when only

the slowest modes (phonons with low energy $h\nu$ and long wavelengths) take part in thermal motion, this quantitative difference is not present; the specific heats of thread-like and layer structures as well as of cubic lattices are proportional to $(T/\theta)^3$ according to equation (105). However, the presence of chain and layer structures manifests itself clearly in the temperature dependence of specific heat at higher temperatures. Tarasov has worked out the theory of the specific heats of such structures; he has obtained very good agreement between the theoretically predicted and experimentally observed behaviour for some substances, including graphite.[8]

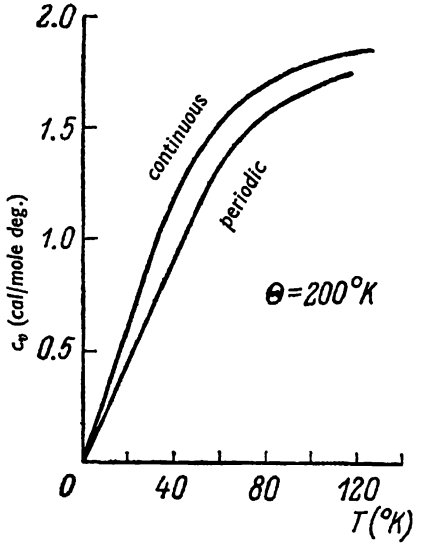


Fig. 75. Temperature dependence of the specific heat for a one-dimensional lattice.

Scattering of electrons

In a perfect crystal, possessing a strictly periodic electrostatic potential at the absolute zero, free electrons and holes would move as in a vacuum. The influence of the medium would manifest itself by the substitution of some effective mass m^* for the free electron mass m_0. In this case there could be no question of defining the mobility as the ratio of the average drift velocity Δv of the charge carriers to the electric field $\mathscr{E}$. A potential difference V applied to such as crystal would accelerate the carriers imparting to them a kinetic energy

$$\tfrac{1}{2}m^* v^2 = eV = e\mathscr{E}L$$

where L is the distance between the electrodes on the crystal.

From this it follows that

$$v = \sqrt{\left(\frac{2e\mathscr{E}L}{m^*}\right)}.$$

Were we to insist on defining a mobility u we would find that u depends upon the dimension L and the field strength $\mathscr{E}$,

$$u = \frac{v}{\mathscr{E}} = \sqrt{\left(\frac{2eL}{m^*\mathscr{E}}\right)}.$$

The fact that the mobility u is independent of the dimensions and the field derives from the fact that the regular acceleration of electrons is continually disturbed by deviations from perfect periodicity of the structure of real crystals. These imperfections may be of various types: cracks, the displacement and twist of individual regions, fluctuations of the thermal motion, the presence of impurities—atoms or ions—or, conversely, the absence of some atoms or ions from their lattice sites. Each of these obstructions disturbs the electronic motion by scattering the electronic waves or, in other words, by creating a high probability of an appreciable deviation of the electronic path. A large number of such imperfections distributed randomly throughout the crystal causes the electrons to return to their equilibrium states, corresponding to a given temperature, by deflecting them in all directions and exchanging energy with them.

Two processes are, thus, seen to be continuously in opposition to each other: (1) the influence of the electric field which displaces the equilibrium towards the predominance of velocities in the direction of the forces acting on the electrons and (2) the scattering influence of the imperfections. These two influences result in a new distribution of electrons in quantum states which departs from the initial distribution the more strongly as the field is increased and the disturbances of the regular motion become comparatively less frequent.

The new distribution function f of the electrons in the quantum states may be represented as a superposition of the normal function f_0 and a small additional function Δf.

The departure from the function f_0 due to the electric field is shown schematically by the dotted line in Fig. 19. The number of states corresponding to velocities directed against the force of the field $\mathscr{E}$ is reduced while the number of states with velocities in the direction of the force $\mathscr{E}$ increases.

The magnitude of the displacement of the new distribution from f_0 in a given field $\mathscr{E}$ depends upon the rate at which the scattering centres restore the disturbed equilibrium. It is natural to assume that, in the case of small departures from the equilibrium distribution, the rate

$d\Delta f/dt$ is proportional to the magnitude of the departures $(f-f_0)$

$$\frac{d\Delta f}{dt} = \frac{df}{dt} = -\frac{f-f_0}{\tau} \tag{108}$$

where τ is a constant characteristic of the rate of the restoring process.

Thus: $$f-f_0 = (\Delta f)_{t=0} \exp[-(t/\tau)]. \tag{108a}$$

The quantity τ is called the relaxation time.

The motion of the electrons may be described very successfully in terms of the concepts of the kinetic theory of gases. By analogy with this theory it is possible to consider the electron as moving continuously for a certain time in a straight line with an acceleration imparted to it by the electric field. On traversing some distance l, known as the free path, the electron meets a lattice imperfection, which disturbs its regular motion and throws it in a different direction. Seen from the point of view of the wave concept, this corresponds to the scattering of plane electronic waves which represent linear motion with constant velocity. Scattering of the waves gives rise to a distribution of probabilities of electronic motion in new directions. Electrons in a crystal lattice are subject to quantum laws governing the changes of momentum and energy. In the presence of the scattering by thermal vibrations, energy may, thus, be taken from, or transfered to, the thermal motion of the lattice in the form of phonons with an energy $h\nu$, while the wave vector may change by $2\pi/a$.

This model presupposes that imperfections are distributed in the crystal at distances which are large compared with the length of the electronic waves and with the dimensions of the imperfections themselves. Only under these conditions does the electron reach a definite velocity both before and after a collision.

We shall assume, moreover, that the imperfections under consideration are sufficiently pronounced so that an electron in collision with one of them has an equal probability of recoiling in any direction, irrespective of the direction of the motion prior to collision.

If the electron has a velocity v the duration of its free motion on a path l is τ' equal to l/v. At the end of this time, moving with acceleration $e\mathscr{E}/m^*$, the electron will have increased the component of its velocity in the direction of the field by $e\mathscr{E}\tau'/m^*$. The mean excess velocity in a unit field is called the mobility and is given by

$$u = \Delta v = \frac{1}{2}\frac{e}{m^*}\tau' \tag{109}$$

and, hence, the current density due to a unit field, that is the electrical

conductivity σ, is:

$$\sigma = neu = \frac{1}{2}\frac{e^2}{m^*} n\bar{\tau}' \tag{110}$$

where n is the concentration of electrons in the conductor.

The computation of the mean value $\bar{\tau}'$ presents considerable difficulties. Furthermore, the very assumption that each scattering centre returns a colliding electron to a completely random state, erasing completely the velocity acquired under the influence of the external field prior to collision, is itself an over-simplification. It is approximately valid only for scattering by foreign impurities and does not apply to thermal vibrations. Although a moving electron encountering a density fluctuation is deflected from its path, the predominant component of its velocity in the new distribution lies in the direction parallel to its velocity prior to collision. The predominance of the initial direction becomes less pronounced after repeated collisions. Finally, after a series of N collisions, no correlation whatever remains between the final and initial directions of velocity. The free path length and the free time between collisions correspond to the sums of the individual free paths and times of flight for all N collisions.

Since, on the average, the component of the velocity in the original direction decreases as a result of numerous statistically distributed deviations, it is preferable to define the time $\bar{\tau}$ as the time during which this component of the velocity decreases by the factor e. In this case it is evident that τ' is equal to τ.

In fact, a very close relation exists between the relaxation time τ and the mean free time of flight τ'. The probability that a collision will occur in a time dt and will erase the change of distribution factor which has arisen before the collision, as a result of the electric field, is, according to equation (108) equal to dt/τ and amounts, in one second, to $1/\tau$. At the same time, if the mean free time of flight of the charge is $\bar{\tau}'$ the number of collisions in one second is $1/\bar{\tau}'$. Thus $\bar{\tau}'$ equals $\bar{\tau}$.

Both ways of describing the return of the disturbed equilibrium to normal lead to the same value of τ, characterizing the rate of the process. The first method using the distribution function is much more general and is not restricted to any particular scattering mechanism. The second method does not require the consideration of all the individual scattering events which combine in the process of re-establishing the normal state but in some cases this method is completely inapplicable.

It has been stated that the thermal energy changes only by a very small amount under the action of an electric field along a path l. With l equal to 10^{-7} cm and a field of 10^2 V/cm the increment $e\mathscr{E}l$ is 10^{-5} eV, while the mean energy of thermal motion is kT or about 0·03 eV.

In some cases, however, the ratio between $e\mathscr{E}l$ and kT may be quite different. Thus, at low temperatures, the mobility of electrons in germanium reaches some hundreds of thousands cm²/V sec. An electric field of 10 V/cm acting over a short time would cause an excess velocity Δv of some millions cm/sec, while the average thermal velocity of non-degenerate electrons at 100° K is $6{\cdot}5 \times 10^6$ cm/sec, i.e. is of the same order as Δv.

In atomic crystals, such as germanium, an increase of the kinetic energy ϵ of the electrons due to the electric field does not alter the mean free path l and, thus, has little influence on the magnitude of the mobility u. In ionic crystals, on the other hand, l is proportional to ϵ and an appreciable increase of velocity in comparison with the thermal velocity leads to substantial changes of both l and u. Thus, Rekalova produced fields of the order $(1\text{–}2) \times 10^6$ V/cm in thin films of As_2S_3. With a mean free path for thermal electrons of 10^{-7} cm this implies an increase of energy of 0·1–0·2 eV while the average thermal energy is only 0·03 eV. As a result the mean free path increases and so does the amount of extra accumulated energy. The effective mean free path appears substantially higher than the value obtaining at low fields; thus the mobility increases with field, as found experimentally by Rekalova.[9]

The thermal fluctuations of the crystal lattice are an essential source of scattering of electronic waves in all conductors. In the strictly periodic field of a perfect lattice, electrons would move completely freely and, in a constant electric field $\mathscr{E}$, they would have a uniform acceleration $\mathscr{E}e/m^*$, where m^* is the effective mass.

Any disturbance of the periodicity of the lattice distorts the electronic path by establishing the probability of a deflection, sometimes causing a change of the electronic energy $\frac{1}{2}m^*v^2$ by $\pm h\nu$ (where ν is the frequency of oscillation of the phonon) with a corresponding change of momentum.

In this case, the laws of conservation of energy and momentum must be satisfied simultaneously:

$$\frac{h^2k^2}{2m^*} - \frac{h^2k'^2}{2m^*} = \pm h\nu_\Phi \qquad (111)$$

$$\mathbf{k} - \mathbf{k}' = \pm \mathbf{k}_\Phi$$

where $\mathbf{k}$ and $\mathbf{k}'$ denote the wave vectors of the electron before and after collision and $h\nu_\Phi$ and $\mathbf{k}_\Phi$ are the energy and wave vector, respectively, of the phonon emitted or absorbed as a result of the interaction.

The velocity of propagation of phonons, which is equal to the velocity of sound, has a magnitude of the order 10^5 cm/sec and the velocities of electrons in a semiconductor are equal to about 10^7 cm/sec. Therefore, for the same value of the momentum p equal to $h\nu/v$, the energy of a

phonon is less than that of an electron by the ratio of their velocities v. While the wave vector of an electron changes abruptly by a large amount, its energy changes only very little in comparison with the initial energy. The more frequent the collisions of an electron the lower its mobility u and diffusion coefficient D.

The conditions for scattering of electrons are somewhat different in metals and in semiconductors.

In a metal the energies of the conduction electrons are distributed in a range of a few kT around the Fermi energy. The energy of these electrons is of the order 5 eV while at room temperature kT is less than 0·03 eV. The spread of velocities of individual electrons amounts in this case altogether to a few per cent. As a consequence we may regard all electron velocities, both in metals and in strongly degenerate semiconductors (in which the room temperature concentration of current carriers appreciably exceeds $2{\cdot}5 \times 10^{19}$) as equal and temperature independent.

In semiconductors to which classical statistics are applicable as a good approximation the spread of energies is, as in metals, of the order kT. As the mean energy is of the same order kT, however, so that the spread is not small in comparison with the mean energy, it is necessary to consider the difference between the velocities of individual electrons. The relaxation time τ, for a given free path l, is then found to be different for electrons or holes with different kinetic energies.

Another important difference exists between the scattering of electrons in metals and semiconductors. While the conduction electrons in metals have energies of the order 5 eV and a wavelength λ of about 5×10^{-8} cm, the mean thermal energy of electrons in semiconductors amounts to only 0·03 eV, with a wavelength of the order 7×10^{-7} cm.

Of all the inhomogeneities present in a crystal, the most effective in scattering the electronic waves are those with dimensions approaching $\lambda/4$. Let us replace an inhomogeneity in the crystal by a plate of thickness $\lambda/4$. The electronic waves reflected from the front and back faces of the region have a path difference $\lambda/2$ and a phase difference π to which is added a further phase difference π on reflection from one of the faces. Thus the total phase difference of waves reflected from the front and back faces equals 2π. Both reflected waves, being in phase, reinforce each other. This simplified reasoning is applicable to the case of normal incidence of waves on the plane, parallel plate. Irregularity of shape of the inhomogeneities and different angles of incidence alter quantitatively the requirements underlying the dimensions for maximum reflection but the order of magnitude remains the same.

Inhomogeneities of dimensions exceeding $\lambda/4$ also cause considerable reflections. On the other hand, imperfections of dimensions small

compared with $\lambda/4$ do not disturb the flow of electronic waves around them appreciably and consequently contribute a negligible amount of scattering.

For these reasons electrons in metals are strongly scattered by all irregularities, beginning with those of atomic dimensions, while in semiconductors only aggregates extending over a number of interatomic dimensions may be expected to contribute actively to the scattering. It follows, for example, that the scattering of electrons in liquid metals is determined by differences of mutual orientation and interatomic distances. On the other hand, electrons in semiconductors, which are scattered by much coarser imperfections, may suffer relatively weak scattering in the liquid state. For optical waves, with wavelengths greater by a factor 100, even these irregularities are insignificant. Thus, amorphous glass and liquid water scatter light only slightly more strongly than crystalline quartz or rock salt.

The fluctuations in the thermal motion always have the same form and dimensions but their number is the larger, the shorter their wavelength. One may, therefore, expect that the scattering of electrons in semiconductors, which is only caused by the coarser imperfections, will be weaker than in metals.

All these considerations are confirmed by experiment. Thus the highest mobilities of some thousands cm^2/V sec at room temperature are found for semiconductors having an atomic lattice (Ge, Si, HgSe) with practically no optical vibrations. In polar semiconductors the mobility attains values of some tens of units. In crystals of Se, TiO_2 or WO_3 the mobility is less than unity.

The different conditions of scattering prevailing in the different types of crystal influence not only the magnitude of the mobility but, also, its temperature dependence.

Mean free path and mobility of electrons

The mean free path of electrons in a metal may be expressed as a function of their energy ϵ and the temperature T:

$$l \propto \frac{\epsilon^2}{T}. \tag{112}$$

In semiconductors with an atomic lattice scattering of electrons is due primarily to longitudinal acoustic vibrations of the lattice which give rise to local compressions and rarefactions and, hence, to fluctuations of the dielectric constant. The most effective of these are the vibrations having wavelengths of an order of magnitude equal to or greater than one quarter of the electronic wavelength (7×10^{-7} cm at room temperature). The free path length in such lattices does not depend

upon the velocity of electrons and is inversely proportional to the temperature

$$l \propto \frac{1}{m^{*2} T} \tag{113}$$

and the mobility may be represented in the form:

$$u = \frac{4le}{3\sqrt{(2\pi m^* kT)}}. \tag{114}$$

Taking into account the dependence of l upon T we obtain:

$$u \propto \frac{C^2}{m^{*\frac{5}{2}} T^{\frac{1}{2}}} \tag{115}$$

where C^2 is a constant of interaction which differs from one material to another.

In semiconductors with ionic lattices the conditions of scattering for electrons differ according to the range of temperature under measurement. At low temperatures, where the energy $h\nu_0$ of the most rapid longitudinal optical lattice vibrations is large in terms of kT, Davydov and Shmushkevitch[10] have shown that

$$l = \frac{a^3}{2\pi} \frac{M}{m^*} \left(\frac{h\nu_0}{\gamma Ze^2}\right)^2 \exp\left(\frac{h\nu_0}{kT}\right) \sqrt{\left(\frac{\epsilon}{h\nu_0}\right)} \tag{116}$$

where M is the ionic mass, a the lattice constant, Ze the ionic charge and γ a coefficient determined by the polarizability of the ions. The free path in such lattices is

$$l \propto \frac{v}{m^*} \exp\left(\frac{h\nu_0}{kT}\right) \tag{117}$$

and the mobility in this temperature range is:

$$u = \frac{a^3}{2\sqrt{(2\pi e^3)}} \frac{M}{m^{*\frac{5}{2}}} \frac{(h\nu_0)^{\frac{1}{2}}}{\gamma^2 Z^2} \exp\left(\frac{h\nu_0}{kT}\right). \tag{118}$$

The dependence of the mobility upon the temperature and the effective mass in the temperature range under consideration, i.e. below the Debye temperature θ, is given by:

$$u \propto m^{*-\frac{5}{2}} \exp\left(\frac{h\nu}{kT}\right). \tag{118a}$$

In the region of high temperatures where $T \gg \theta$ and $h\nu_0 \ll kT$ the free path is determined by the following relation:

$$l = \frac{a^3}{2\pi} \frac{M}{m^*} \left(\frac{h\nu_0}{\gamma Ze^2}\right)^2 \frac{\epsilon}{kT}. \tag{119}$$

Thus, in this case: $$l \propto \frac{v^2}{m^* T}. \tag{120}$$

The electronic mobility in ionic semiconductors at temperatures sufficiently higher than the Debye temperature is

$$u = \frac{4a^3}{2\pi e^3} \frac{M}{m^{*\frac{3}{2}}} \left(\frac{h\nu_0}{\gamma Z}\right)^2 \frac{1}{\sqrt{(2\pi kT)}} \tag{121}$$

from which we obtain the following dependence of mobility u upon temperature T and effective mass m^*:

$$u \propto m^{*-\frac{3}{2}} T^{-\frac{1}{2}}. \tag{122}$$

It is found that all these types of scattering by thermal vibrations introduce changes into the distribution functions of the electronic quantum states, which lead to values of the mobility increasing with decreasing effective mass of the current carriers and decreasing with increasing temperature.[11]

The scattering of holes and electrons by positive or negative ions has a different character. In ionic lattices, thermal motion and diffusion always give rise to vacant sites abandoned by one of the lattice ions, or excess ions diffusing between the lattice cells or, finally, to the presence of foreign ions of a valency different from that of an atom in the host lattice. In atomic lattices, to every electron or hole there corresponds an ionized impurity atom which gives off or, conversely, accepts an electron. In each of these cases the ion is surrounded by an electrostatic coulomb field, which is weaker than that in a vacuum by a factor ϵ, where ϵ is the dielectric constant of the lattice. An electron moving in this field is deflected from its initial path the more strongly, the closer it approaches the ion and the longer it remains in the field, i.e. the slower its motion.[12, 13]

The duration of free flight between successive deflection by ions having a concentration N is given by:

$$\tau = \frac{8\sqrt{2}}{\pi^{\frac{1}{2}}} \frac{m^{*\frac{1}{2}} \epsilon^2}{Ne^4} \frac{(kT)^{\frac{3}{2}}}{\ln[1+(3\epsilon kT/e^2 N^{\frac{1}{3}})^2]}$$

and the mobility:

$$u = \frac{e}{m^*}\tau = \frac{8\sqrt{2}}{\pi^{\frac{1}{2}}} \frac{\epsilon^2}{Ne^3 m^{*\frac{1}{2}}} \frac{(kT)^{\frac{3}{2}}}{\ln[1+(3\epsilon kT/e^2 N^{\frac{1}{3}})^2]}. \tag{123}$$

Neglecting the relatively weak temperature dependence introduced by the logarithmic term in the denominator, the mobility for the case of ionic scattering may be represented in the following form:

$$u_i \propto \frac{T^{\frac{3}{2}}}{Nm^{*\frac{1}{2}}}. \tag{124}$$

In the case of a degenerate electron gas in a semiconductor (where the concentration of electrons and impurity ions at room temperature is much greater than $2{\cdot}5\times10^{19}$ cm^{-3}) the formula for ionic scattering becomes:

$$u_i = \frac{3h^3\epsilon^2}{16\pi^2 e^3 m^*[\ln(1+x)-x/(1+x)]} \tag{124a}$$

where

$$x = \left(\frac{h}{e}\right)^2 \frac{e}{m^*}\left(\frac{3N}{8\pi}\right)^{\frac{1}{3}}.$$

The relation given by expression (124) may easily be understood qualitatively. Firstly, each of the N ions introduces its own contribution to the resistance independently of the remaining $(N-1)$ ions. Therefore, N ions lower the mobility N times more than does one ion. Secondly, the higher the effective mass of a carrier the lower its velocity at a given temperature T, since $\frac{1}{2}m^*v^2$ equals $\frac{3}{2}kT$; hence the more marked is its deflection when travelling at a given distance from the ion. Thirdly, the mobility increases with the velocity as the temperature increases. The magnitude of the deflection is an inverse function of the velocity of the electron passing at a given distance from the ion. The bigger the individual deflection, the smaller the number of such deflections required to scatter the carriers through all possible angles up to 2π.

Since the velocity v equals $(3kT/m^*)^{\frac{1}{2}}$, we have, according to equation (124),

$$u_i \propto \frac{m^*}{N}v^3 \propto \frac{T^{\frac{3}{2}}}{Nm^{*\frac{1}{2}}}. \tag{125}$$

In practice we find scattering both by the thermal vibrations and by the impurity ions. The higher the temperature and the purer the semiconductor, the more important the thermal vibrations and the less important the scattering by ionized impurities.

In the general case we may take into account both mechanisms of scattering with sufficient accuracy by assuming that the number of collisions suffered by an electron in 1 cm of its path is equal to the sum of the number of collisions due to thermal vibrations together with the number of deflections by the Coulomb fields of ions. The free path l is inversely proportional to the number of collisions suffered by an electron in 1 cm of its path. These collisions are responsible for maintaining equilibrium distribution of velocities. Consequently

$$\frac{1}{l} = \frac{1}{l_i} + \frac{1}{l_T}$$

or, approximately:

$$\frac{1}{u} = \frac{1}{u_i} + \frac{1}{u_T}. \tag{126}$$

For the case of atomic lattices, equation (126) leads to the following relation:

$$\frac{1}{u} = aT^{-\frac{1}{2}} + bT^{\frac{3}{2}}. \qquad (127)$$

It would be possible to extend our considerations to the mechanisms of scattering by neutral impurity atoms, and by dislocations caused by plastic deformation of the crystal. It can easily be shown, however, that neither of these processes can have an appreciable influence on the magnitude of the mobility.

In fact, the scattering cross-section of a neutral atom is of the order of $(3 \times 10^{-8})^2$ or about 10^{-15} cm^2. With an impurity concentration of 10^{19} cm^{-3} the total scattering cross-section amounts to approximately 10^4 cm^2, while the scattering cross-section of ions of a similar concentration is of the order 10^6 cm^2, and that of thermal vibrations is higher still, about 2×10^7 cm^2.

Plastic dislocations have an even smaller effect. With dislocations spaced at a distance of 1 μ apart, the total scattering cross-section is 10^4 cm^2, while in the majority of cases it is much less.

Ansel'm has shown, however, that scattering by neutral atoms at low temperatures may be considerably stronger than indicated by the above calculations, especially if the impurity atoms give rise to energy levels very close to the conduction band (even if not occupied by electrons). In these circumstances a slowly moving electron may be scattered even at considerable distances from the centre of the atom, as much as 10 times farther away than the figure assumed in the foregoing calculations. In this case the scattering cross-section is increased by a factor 100 and this scattering mechanism may become more important than scattering by the thermal vibrations. Ansel'm suggested the name of resonance scattering for this mechanism of interaction between the free electrons and the impurity atoms.[14]

The free path length of the current carriers may be determined not only from the magnitude of mobility but, also, in a more direct way. The free path, which at room temperature has an order of magnitude $10^{-7} - 10^{-6}$ cm, increases at lower temperatures and at liquid hydrogen temperature reaches the magnitude of several microns; a still higher value holds at liquid helium temperature. If the transverse dimensions of the specimen are less than the free path length, the majority of collisions responsible for scattering will occur at the external boundaries leading to a reduction of the electrical conductivity.

A measurement of the dependence of the resistivity upon the specimen thickness for thin wires or films may enable a determination of the magnitude of the free path length l to be made. However, in the case of thin films, which may be easily obtained both with metals and

semiconductors by vacuum evaporation, uncertainties arise with regard to their continuity and uniformity. An increase of sensitivity could, therefore, be ascribed to the appearance of contact resistances between grains forming the film.

Much more conclusive are results obtained with thin wires. In the case of metals, the experimental results thus obtained agree well with theoretically predicted values. It has proved to be impossible, so far, to prepare sufficiently thin wires from semiconductors.

Another method for estimating the free path length of the current carriers makes use of measurements of resistivity at sufficiently high frequencies for an electromagnetic field to penetrate only a very small distance into the material (skin effect).

Finally, an estimate of the free path length may be made by measuring the effect of a transverse magnetic field on the electrical conductivity. If the radius of the electronic orbits in a magnetic field becomes comparable with the free path, the angles of deflection of electronic paths may reach 90° and the character of the dependence of resistivity upon the magnetic field may be changed. It may be seen that, with deflection through 180°, an electron current, which in weak magnetic fields, or with short free paths, would be deflected to the left, will be deflected to the right.

Unfortunately, there are no direct data relating to the magnitude of the free path length in semiconductors.

A comparison between the formulae quoted above and experimental results for the temperature dependence of the conductivity gives reasonably good agreement only in the case of atomic lattices, for which the main source of scattering is due to the acoustic vibrations. Here the relation $u = u_0 T^{-\frac{3}{2}}$ is approximately verified. Thus, for example, in Ge the mobility of electrons follows the relation $u = u_0 T^{-1\cdot65}$, while for holes $u = u_0 T^{-2\cdot3}$.

The relation $u = u_0 T^{-\frac{1}{2}}$, for $T \gg \theta$, is very frequently not satisfied for semiconductors in which the optical vibrations play an important role. For crystals such as Cu_2O and PbS Pekar obtained good agreement with the formula for scattering by polarons which may be written, for temperatures which are not excessively high, in the form

$$u = u_0\left[\exp\left(\frac{h\nu_0}{kT}\right) - 1\right]$$

where ν_0 is the limiting vibrational frequency of the ions in the lattice.[15]

For Cu_2O a value for $h\nu_0$ of 0·031 eV was obtained, while for PbS $h\nu_0$ is equal to 0·074 eV and u_0 equals 14 cm²/V sec. These values of ν_0 do not agree, however, with the experimental values of the Debye temperature θ which is related to ν_0 by the equation $h\nu_0 = k\theta$. If, however, the mobility in PbS is represented in the form $u = u_0 T^{-n}$, the

exponent n increases continuously when passing from low to high temperatures.

In the case of some sulphides and tellurides, measurements of the mobility over a wide range of temperatures—from -100 to $+400°$ C—agree well with the formulae $u = u_0 T^{-3}$ or $u = u_0 T^{-\frac{3}{2}}$. In the case of MnTe the exponent of T amounts to between $-\frac{5}{2}$ and $-\frac{7}{2}$, i.e. approximately to -3. Fig. 77 shows the dependence of $\log u$ upon $\log T$ for different values of the exponent.

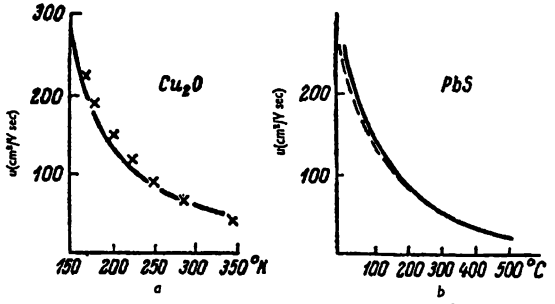


Fig. 76. Temperature dependence of the carrier mobility in Cu_2O and PbS: (a) the continuous curve corresponds to $u = 27$ [exp $(363/T) - 1$] and the crosses refer to the experimental data of Engelhard; (b) the continuous curve corresponds to $u = 14{\cdot}7$ [exp $(865/T) - 1$] and the dotted line gives the experimental results of Dunaev and Maslakovets.

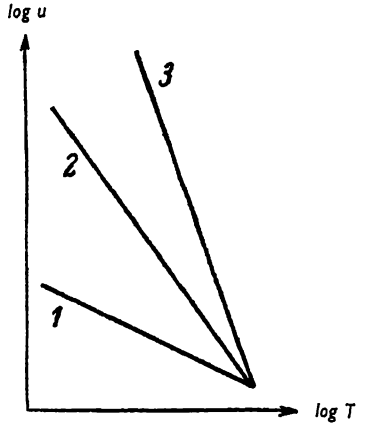


Fig. 77. Dependence of carrier mobility on temperature: (1) $u = A_1 T^{-\frac{1}{2}}$; (2) $u = A_2 T^{-\frac{3}{2}}$; (3) $u = A_3 T^{-3}$.

Peierls and Samoilovich have pointed to one more further circumstance, which influences the mobility of holes and electrons in semiconductors in which the free path depends upon the electronic energy ϵ.

The uncertainty principle gives the following relation between the uncertainty $\Delta\epsilon$ in the value of energy and the duration τ of the relaxation process which is determined by this energy:

$$\Delta\epsilon = \frac{h}{\tau}.$$

The average duration of motion of an electron with velocity v over a distance l, or the relaxation time is:

$$\tau = \frac{l}{v}.$$

At the same time the electronic wavelength is determined by the expression

$$\lambda = \frac{h}{m^* v}.$$

Consequently:

$$\Delta\epsilon = \frac{hv}{l} = 2\frac{\lambda}{l}\epsilon$$

and

$$\frac{\Delta\epsilon}{\epsilon} = 2\frac{\lambda}{l}.$$

At room temperature λ is 7×10^{-7} cm. Thus, if the free path $l < 10^{-6}$ cm, the uncertainty in the value of the energy exceeds the value of energy itself and equations (118) and (121), which have been derived formally without taking into account the uncertainty principle, lose their validity.

While the uncertainty of the value of kinetic energy of an electron does not play any appreciable part in the determination of the free path length in atomic lattices, where l does not depend upon ϵ, in ionic lattices, in which l is proportional to ϵ, such values of $\Delta\epsilon$ which exceed ϵ by a factor of 10 are completely inadmissible.

In the case of ionic lattices, in which the optical vibrations cause a marked scattering of the current carriers, calculation leads to values of the free path l considerably less than 10^{-7} cm and even less than 3×10^{-8} cm, i.e. less than the interatomic distance. This not only casts doubts upon the results of the analysis, but deprives the basic assumptions of the theory relating to the free passage of an electron over distances of the order l of physical meaning. The application of classical laws of motion and the definition of a velocity of motion $\bar{v}$ are only permissible for motion averaged over several interatomic distances. Within the range of a single lattice cell the movement of an electron is extremely complicated and its path cannot possibly be described by a definite trajectory over distances comparable with the extension of the

wave function. The free path length l in such cases is found to be much shorter than the electronic wavelength.

Apart from this, if the mean free path between two scattering events is l, the geometrical extension of inhomogeneities which cause this scattering should be considerably less than l, otherwise the whole volume of the crystal would appear to be filled with these inhomogeneities. There would be no room left for "free" movement in an undistorted periodic lattice.

In ionic lattices λ exceeds l by almost an order of magnitude and, thus, the electronic wavelength λ ought to exceed by ten times the dimensions of inhomogeneities encountered in its path. However, in this case the scattering cannot have the character which we have ascribed to it and which underlies the identification of the mean time of free flight with the time of relaxation.

In principle, electronic waves flow round any inhomogeneities which are small in comparison with their wavelength. Not all directions, by any means, are equally probable after scattering. On the contrary, after scattering, the velocity distribution is likely to retain a considerable component parallel to the velocity prior to scattering.[16]

From whatever angle we approach the process of scattering of electrons in a semiconductor which has a low mobility, we are invariably confronted with this insurmountable difficulty—the inapplicability of the concepts of a free path, or even of a mean velocity and energy, of an electron.

Thus, we find that methods applicable to the study of metals and atomic lattices break down in the case of ionic lattices. No wonder, therefore, that the results of these methods are in disagreement with experimental facts. The discrepancy between experimental and theoretical data is illustrated in Fig. 77. The relation $uT^p = \text{const.}$ is represented in a double-logarithmic graph by a straight line with a slope equal to the exponent p.

For the case of ionic lattices we should expect $p = \frac{1}{2}$ (line 1), for atomic lattices $p = \frac{3}{2}$ (line 2), whilst measurements on a number of semiconductors with ionic binding give lines with much steeper slopes, $p = 3$ (line 3) and even $p > 3$. Such a strong temperature dependence of the mobility cannot be explained by any of the known types of binding. It may be noted, however, that a direct relation exists between the values of the exponent p and the dependence of the free path length upon electronic velocity.

Attempts have been made to explain the strong temperature dependence of the mobility starting with the assumption of multi-quantum transitions during recombination. Alternatively it has been assumed that electrons interact, not with low-energy phonons, the

spectral density of which increases with increasing momentum, but with high-frequency phonons, the number of which depends very little upon the frequency. However, it is difficult to justify these assumptions.

The mobility of electrons is closely related to their diffusion under the influence of a concentration gradient dn/dx in the semiconductor. Einstein established the following general relation between the mobility u of current carriers and their diffusion coefficient D_e:

$$\frac{u}{D_e} = \frac{e}{kT}. \tag{128}$$

This relation has been confirmed to a high degree of precision for semiconductors. If an increased concentration of carriers of one type is established in any part of a semiconductor, their diffusion charges the surrounding regions, while the parts which the carriers have left acquire charges of the opposite sign. An electric field $\mathscr{E}$ is established between the part with the higher concentration and the neighbouring regions, and opposes the motion of the diffusing charges. In a very short time equilibrium is established satisfying the equation:

$$D_e \frac{dn}{dx} = nu\mathscr{E} \tag{129}$$

where D_e is measured in cm²/sec, n in cm^{-3}, x in cm, u in cm²/V sec and $\mathscr{E}$ in V/cm.

In the case of a non-uniform distribution of electrons in an electric field $\mathscr{E}$, the transport of charges in each volume element of the material is expressed as the difference between the drift and diffusion currents:

$$j = nu\mathscr{E} - D_e \frac{dn}{dx}. \tag{129a}$$

The conditions of diffusion are changed drastically by the presence of current carriers of a different sign. We shall assume, at first, that both the concentrations of electrons and of holes are equal to n. If their diffusion coefficients, and, hence, also their mobilities, are not equal, the faster carriers move ahead and cause a space charge. This accelerates the slower carriers of the opposite sign, while slowing down the faster carriers. From this time onwards both species of free carrier will move with a common velocity and an electric field $\mathscr{E}$ will be established in the solid, in the direction of the concentration gradient, equalizing the velocities of positive and negative charges. The free charge carriers diffuse unhindered in this field in the direction of the concentration gradient. If the mobilities of both carrier species are identical, the field is zero, but, even in the case of different mobilities, the field $\mathscr{E}$ is small—

much smaller than the field which would be necessary to prevent diffusion altogether in the presence of one type of carrier.

The diffusion of the free carriers should be considered independently from their mobility in the electric field due to the diffusion process itself.

Consider a region of increased charge concentration in a limited region of a semiconductor, e.g. near a point contact electrode. This charge density diffuses outwards into the surrounding medium only if both types of carrier are generated simultaneously or if the point contact introduces carriers of the opposite type to that predominating in the absence of injection. If, on the other hand, the point contact only introduces the type of carrier which originally determined the current mechanism, diffusion will be opposed by the formation of a charge double layer at a small distance away from the initial region.

In the presence of n electrons and p holes per cm^3 in a semiconductor, the diffusion coefficient of their common current is

$$D = \frac{n+p}{(n/D_P)+(p/D_N)}$$

where D_P and D_N are the diffusion coefficients for holes and electrons separately. In the case where $n = p$ the ambipolar diffusion is determined by the coefficient

$$D = 2\frac{D_N D_P}{D_N+D_P}.$$

At least two more mechanisms of carrier motion may exist in a semiconductor.

One of these arises when the impurity density is so high that the electronic wave functions of impurity atoms begin to overlap and form a new band only partly filled with electrons. The width of such an impurity band is much less than that of the conduction band, both as a result of much weaker overlap of wave functions and much weaker electrical interaction between impurity atoms, and, also, in view of the fact that the density N of impurity atoms is less than the total density N_0 of lattice atoms by one or two orders of magnitude. Thus, the density of levels in the impurity band is less than that in the conduction band in the ratio N/N_0. The conduction mechanism in an impurity band is likely to be of a metallic nature.[17]

The appearance of overlapping wave functions is favoured by a high dielectric constant. If the extensions of wave functions of the valence electrons are insufficient for the formation of an impurity band, the band may still be formed by the excited levels of the impurity electrons, the spatial extension of which is 4 times larger. The band of excited levels of impurity atoms may be energetically nearer the ground levels of the impurity atoms than the conduction band. In this case electrons

will make transitions into the impurity band in larger numbers than into the conduction band. The small width of the impurity band will, however, cause a considerable lowering of the mobility.

This type of conduction may become important at low temperatures when the carrier concentration in the conduction band is low. Experimental data which could be interpreted in terms of the appearance of an impurity band have given values of the mobility between 10^{-4} and 100 cm^2/V sec, depending upon the concentration. The mobility in the conduction band of the crystal of these temperatures is of the order 10^5 cm^2/V sec.

The second mechanism of mobility has been studied by Verwey for various compounds of the transition elements. As a simple example, we may take the compound Fe_3O_4 in which it is possible to assume the existence of both divalent and trivalent iron ions, arranged alternately in rows along certain crystallographic directions. Under the influence of thermal motion some of these ions exchange places; a trivalent ion is replaced by a divalent and vice versa. Such a transition does not alter the over-all energy, but may require some activation energy ΔE. If an activation energy must be expended, the probability of transition, and, hence, also the velocity of diffusion of electrons, tends to increase with temperature as $\exp(-\Delta E/kT)$.

In an external electric field $\mathscr{E}$ the probability of transition in one direction increases, and that in the opposite direction decreases. As a result, an electric current appears in the direction of the electric field and is proportional to its magnitude.

This mechanism does not require the preliminary transition of electrons from a normal state into free state and the quantity ΔE has a special physical meaning.

It is difficult to apply the concept of a free path to this type of conduction, but the concept of mobility still retains its meaning.

It is thought that this conduction mechanism is present in a large class of semiconductors, known as inverse spinels and possessing pronounced ferromagnetic properties.[18, 19]

The concentration of charge carriers

The formulae determining the equilibrium densities of holes and electrons indicate that their concentrations depend upon the position of the chemical potential level relative to the edge of the conduction or valence band:

$$n = \frac{2(2\pi m^* kT)^{\frac{3}{2}}}{h^3} \exp\left(-\frac{\Delta E'}{kT}\right) = N_{\text{eff}} \exp\left(-\frac{\Delta E'}{kT}\right) \qquad (130)$$

where $\Delta E'$ is the distance of the chemical potential level from the edge

of the band. In particular for intrinsic conduction, assuming equal effective masses of holes and electrons, we have $\Delta E'$ equal to $\Delta E_0/2$, where ΔE_0 is the width of the forbidden band.

Fig. 78 shows the relation between the intrinsic concentration of holes and electrons n_0 and the width of the forbidden band ΔE_0. The abscissae give the energy ΔE_0 and the ordinates the logarithm of the concentration. The graph is based on the assumptions that $m_+^* = m_-^* = m_0$. It is evident that for other values of m^* a term $\ln (m^*/m_0)^{\frac{3}{2}}$ must be added to $\ln n$.

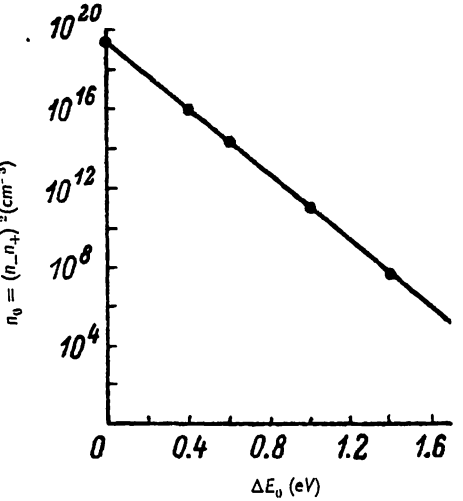


Fig. 78. Relation between the intrinsic carrier concentration and the width of the forbidden gap.

In the general case of combined impurity and intrinsic conduction the graph gives the magnitude of $n_0 = \sqrt{(n_+ n_-)}$.

The most common method of determining the concentrations experimentally consists in the measurement of the Hall effect. This method is relatively simple if we are certain that only one type of current carrier is present. The Hall constant in this case is given by the relation:

$$R = \frac{A}{nec}.$$

The coefficient A may have different values, however, according to the type of scattering mechanism. The values of A for various cases are shown in Fig. 103 which gives the dependence of A upon the relative contributions of ionized-impurity and lattice scattering. The maximum variation in the magnitude A is 1·93 to 1, i.e. a factor of approximately 2.

The problem of measuring the concentrations of holes and electrons is still more complicated in the case of mixed conduction. Here it is important to know the density of minority carriers in relation to the density of the majority carriers. It appears that the most sensitive indication of the presence of two types of carriers is represented by the Nernst–Ettingshausen effect. The latter differs from the Hall effect in that the electric current passing through the semiconductor is replaced by a diffusion current of electrons and holes due to a temperature gradient.

Fig. 78 is based on the formula for the concentration n_0 of free carriers in the non-degenerate state

$$n_0 = N_{\text{eff}} \exp(-\Delta E_0/2kT) \tag{131}$$

where N_{eff} is the effective density of states in the band:

$$N_{\text{eff}} = \frac{2(2\pi m^* kT)^{\frac{3}{2}}}{h^3}. \tag{132}$$

The electron gas must be considered to be degenerate if the density of electrons in the conduction band is so high that they fill all levels up to an energy E_Φ, lying appreciably above the edge of the conduction band E_c. Shifrin[20] has established a criterion enabling the classification of electrons in the conduction band as in either the non-degenerate or degenerate state. If the chemical potential or Fermi level falls more than $2kT$ above the edge of the conduction band the electron gas must be considered degenerate. The degeneracy criterion may therefore be written in the form:

$$\frac{\mu - E_c}{kT} > 2.$$

If, however, the level of the chemical potential is situated more than $2kT$ below the edge of the conduction band, the electron gas may be considered non-degenerate. In this case

$$\frac{\mu - E_c}{kT} < -2.$$

When the chemical potential remains within these limits

$$-2 < \frac{\mu - E_c}{kT} < 2$$

the electron gas is in an intermediate state. In the degenerate case, as for metals, the concentration n of electrons in the conduction band is given by the formula

$$n = \frac{8\pi}{3h^3}[2m(\mu - E_c)]^{\frac{3}{2}} = \frac{4N_{\text{eff}}}{3\sqrt{\pi}}\left(\frac{\mu - E_c}{kT}\right)^{\frac{3}{2}}. \tag{133}$$

The concentration of non-degenerate electrons, on the other hand, is given by the expression:

$$n = N_{\text{eff}} \exp\left(-\frac{E_c - \mu}{kT}\right). \tag{130a}$$

Neither equation (130a) nor equation (133) gives the true value of electron concentration in the conduction band for all intermediate cases. Fig. 79 gives the dependence of concentration, in the intermediate range, upon the magnitude of $(\mu - E_c)/kT$. The full line refers to the true value of n. The dotted line on the left denotes the values of n obtained on the assumption of non-degenerate electrons (equation (130a)). The dotted line on the right gives n from equation (133) for completely degenerate electrons.

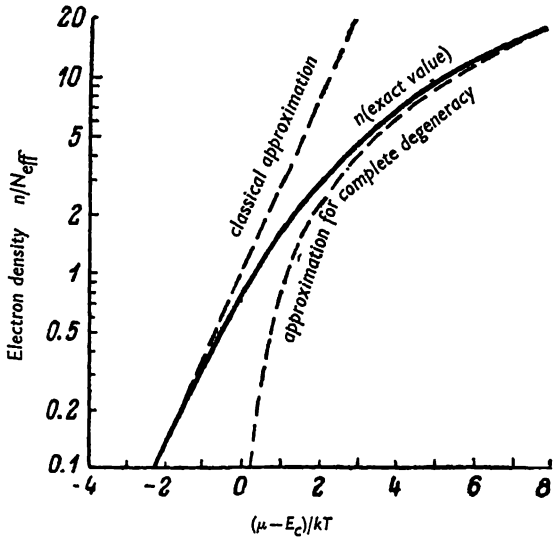


Fig. 79. Carrier concentration as a function of the reduced chemical potential.

Shifrin's criteria may be somewhat modified with reference to Fig. 79. Equation (130) gives reasonably correct results up to $(\mu - E_c)/kT \fallingdotseq -1$. At the same time, in order to obtain similar accuracy in the degenerate case from equation (133) it is necessary to put the condition $(\mu - E_c)/kT > 4$. For $(\mu - E_c) = 0$ equation (130) gives a value of n 20% in excess of the true value. The same error in the sense of a lower value is incurred using equation (133) for $(\mu - E_c)/kT = 2$. In the latter case formula (130) gives a carrier density $2\frac{1}{2}$ times too great.

Properties of surface layers on semiconductors

The boundaries of semiconductors are the seat of a number of phenomena which are present neither in metals nor in insulators. These phenomena have acquired considerable technical importance and have, therefore, been studied in great detail.

Equilibrium between a semiconductor and an adjacent metal corresponds to the equality of the chemical potentials of the current carriers. We have already studied the position of the chemical potential in the system of electronic energy levels of a semiconductor. In a metal the chemical potential coincides with the limit of the Fermi distribution at the absolute zero.

In the case of equality of chemical potentials, the probability of occupation of free electronic levels is the same at any given energy above or below the common chemical potential. The electrons or holes present in these levels move from one medium to the other without altering their total energy, irrespective of the fact that the bonding and potential energies are different on both sides of the boundary. The equality of electrochemical potentials is attained in this case by means of the transfer of electrons or holes from one medium to the other. This charge transfer gives rise to a potential difference which compensates for any differences between the binding energies.

A measure of the binding energy is afforded by the total amount of energy which a charge must lose if passing from a given solid into vacuum. The total is made up of the kinetic energy which the electron possesses at the limit of the Fermi distribution, together with some additional amount of energy. The latter determines the contact potential in the case of a metal. The potential difference established in equilibrium between two bodies is equal to the difference between their contact potentials. A potential step is thus likely to be formed at the interface between two bodies in contact. The surface layers of the two solids thereby become oppositely charged and form a condenser with a surface charge density corresponding to the contact potential difference.

The charge density required for this purpose may easily be estimated.

The contact potential difference is commonly of the order of 1 V, i.e. 1/300 e.s.u. of potential. The spacing between the surfaces in contact is of the order of the interatomic distance, i.e. about 3×10^{-8} cm, so that the charge per cm^2 of the surface is:

$$q = \frac{1}{4\pi \times 3 \times 10^{-8}} \times \frac{1}{300} = 9 \times 10^3 \text{ e.s.u.}$$

The number of charge carriers at the surface is thus:

$$n = \frac{q}{e} = \frac{9 \times 10^3}{5 \times 10^{-10}} \fallingdotseq 2 \times 10^{13} \text{ charges/cm}^2.$$

Taking into account the polarization of the medium, due to the influence of the electric field and depending on the dielectric constant, the density n should be further increased.

In a metal, the electron density is equal to the density of atoms and a square centimetre of a monatomic surface layer contains approximately $(3 \times 10^{-8})^{-2}$, i.e. about 10^{15}, atoms and as many electrons. The contact potential difference between two metals may, therefore, easily be compensated by an excess of electrons on the surface of one metal and a deficit on the other. As a result of this the concentration of electrons in the surface layer changes only by a few per cent.

The potential step between the plates of the condenser, thus formed at the interface, equalizes the Fermi levels in the two metals in contact. As a result of this process the total energy of electrons, at levels equally removed energetically from that of the chemical potential, is equal in both metals. To a first approximation the transition of an electron from one metal into the other may be effected without change of energy. One ought to take into account, however, the differences between the mean velocities of electrons arising as a result of the different kinetic energies in the two metals at the Fermi levels, and of the differences in the electron concentrations. It is found, then, that the transport of electrons from one metal into another is accompanied by the release at the interface of a certain amount of energy (positive for one direction of transport and negative for the opposite direction). In the case of current flow in a closed circuit made of two different metals this gives rise, according to the direction of current flow, to the absorption of Peltier heat at one contact and its release at the other. The amount of heat involved depends upon the temperature.

This is the origin of the thermoelectric phenomena in metals. The order of magnitude of the thermoelectric voltage is a few units or tens of microvolts per degree centigrade temperature difference between the contacts.

Contact resistance

The conditions existing at the interface between a semiconductor and a metal are different from those at the contact between two metals. The concentration of free electrons in a semiconductor is usually 10^{14}–10^{18} per cm^3 and the concentration in a monatomic layer at the surface is $(10^{14}$–$10^{18}) \times 3 \times 10^{-8}$, i.e. 10^{7}–10^{10} electrons per cm^2. Thus, even the whole electronic population of a monatomic surface layer is insufficient for the formation of a potential difference of 1 V.

The potential difference, which is not compensated by charges at the interface, penetrates as an electric field into the bulk of the semiconductor. This causes the accumulation of a space charge required to

equalize the difference between the chemical potentials of the two solids in contact.

In order to calculate the depth L of the space charge layer we shall make the simplifying assumption that the contact field removes all the free carriers initially present in this layer. This converts all the impurity atoms—donors or acceptors—into positive or negative ions respectively with a concentration n equal to n_0 (Fig. 80).

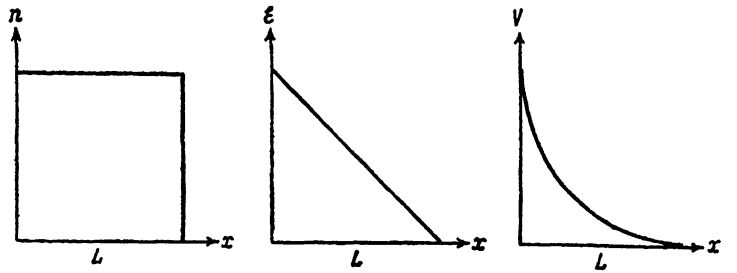


Fig. 80. Distribution of carrier density (n), electric field ($\mathscr{E}$) and potential (V) in a space charge layer (L).

If the concentration of impurity atoms is, say, 10^{20} cm^{-3}, the space charge density is

$$\rho = 10^{20} \times 4 \times 10^{-10} = 5 \times 10^{10} \text{ abs. e.s.u./cm}^3.$$

The increment $d\mathscr{E}$ of the electric field in a thickness dx of space charge density ρ at a distance x from the contact is

$$d\mathscr{E} = \frac{4\pi\rho}{\epsilon} dx.$$

Since deep in the semiconductor the field strength $\mathscr{E}_L$ is zero, the field in the outer layer is given by:

$$\mathscr{E}_x = \int_L^x \frac{4\pi\rho}{\epsilon} dx = -\frac{4\pi\rho}{\epsilon}(L-x). \tag{134}$$

The potential difference V between the inner part of the semiconductor ($x = L$) and the plane x is:

$$V = +\int_L^x \mathscr{E}_x dx = -\frac{4\pi\rho}{\epsilon}\left(Lx - \frac{x^2}{2} - \frac{L^2}{2}\right). \tag{135}$$

The potential difference V_0 between the outer boundary, considered to be the plane of the contact ($x = 0$), and the inner boundary of the space charge layer extending to a depth L is:

$$V_0 = \frac{4\pi\rho}{\epsilon}\frac{L^2}{2}. \tag{136}$$

Assuming a value for V_0 of 1 V, i.e. 1/300 e.s.u., and putting ϵ equal to 10, and $\rho = 5 \times 10^{10}$ e.s.u. cm^{-3}, we obtain:

$$L = \sqrt{\left(\frac{10}{300 \times 3{\cdot}14 \times 2 \times 5 \times 10^{10}}\right)} \doteq 3 \times 10^{-7} \text{ cm.}$$

With an ionized impurity concentration of 10^{16} cm^{-3} we would have L equal to 3×10^{-5} cm.

Variations of the space charge layer thickness L give rise to different qualitative properties of this layer. With $L < 10^{-6}$ cm, the layer thickness may become less than the electronic free path l.

In this case the conditions are analogous to the transition of an electron *in vacuo* from an electrode with a lower work function to an opposite electrode. These transitions take place until the contact field is compensated. With a concentration of ions in the surface layer amounting to 10^{21} cm^{-3} we would obtain the value $L = 3 \times 10^{-8}$ cm and the energy barrier at the surface would be transparent.

We shall consider the case of a layer for which $L \gg l$ (a more accurate condition is $L \gg leV/kT$), corresponding to a low concentration of free carriers in the semiconductor. We shall also renounce the simplifying assumption about the complete ionization of all impurity atoms in the barrier layer. The concentration of free charges will be denoted by n.

At a low concentration of ions the space charge layer extends over several tens of free paths. A definite concentration of free carriers is therefore formed at every point of the layer.

Electrons placed in a field $\mathscr{E}$ drift with the average velocity v equal to $u\mathscr{E}$. The electron current density is:

$$j = nu\mathscr{E}$$

where u is the carrier mobility. At the same time, as the potential V and concentration n of free carriers vary from point to point in the layer, electrons diffuse giving rise to a current

$$j_D = D_e \frac{dn}{dx},$$

where D_e is the diffusion coefficient. Equilibrium in the layer is reached when both currents are equal and opposite, i.e. when at every point in the layer

$$nu\mathscr{E} = D_e \frac{dn}{dx}. \tag{129}$$

The mobility u and the diffusion coefficient D_e are connected by Einstein's relation:

$$u/D_e = e/kT. \tag{128}$$

Apart from this, the following condition must be satisfied at any cross-section

$$(n_+ - n_-)\, e = \frac{1}{4\pi} \frac{d^2 V}{dx^2} \tag{137}$$

where n_+ and n_- are the concentrations of positive and negative charges.

Starting with these assumptions Davydov developed the theory of the surface layer of a semiconductor in contact with a metal. The result depends both on the direction of the contact field within the layer and upon the sign of current carriers.[21]

If the contact potential of the metal is the higher, i.e. its work function for electrons is larger than in the semiconductor, the contact field has the effect of moving the negative electrons present in the surface layer into the interior of the semiconductor. Conversely, holes are drawn into the surface layer. Thus, the surface layer of an n-type semiconductor is depleted of free electrons and becomes filled with a positive space charge. The potential V in this layer is raised in comparison with that in the remaining volume of the semiconductor. In view of the fact that the chemical potential is constant throughout a system in equilibrium, the bottom of the conduction band and all the other energy levels, rise away from the chemical potential level. As we know, this leads to a reduction of the density n of free electrons in the ratio

$$n/n_0 = \exp(-eV/kT).$$

Thus, the surface layer formed in the case of an n-type semiconductor in contact with a metal having a higher work function has a reduced concentration of free charges and an increased resistivity. Conversely the surface layer in a p-type semiconductor draws in holes and builds up a positive space charge by raising the top of the valence band and bringing it nearer to the chemical potential level. The surface layer formed in a p-type semiconductor, in contact with a metal of higher work function, has a higher electrical conductivity.

A similar argument leads to the conclusion that the contact between a semiconductor and a metal with a lower work function has a higher resistance for a p-type semiconductor and higher conductance for an n-type semiconductor.

Fig. 81 shows the positions of the edges of the conduction and valence band in the contact layer for each of the four cases considered. Having investigated a large number of semiconductors and metals and determined their contact potentials and signs of the current carriers, A. V. Ioffe has found that Davydov's theory was applicable in all cases without exception, at least qualitatively. Quantitatively, however, the theory was found to be not sufficiently accurate since it does not take into account the following circumstances.[22]

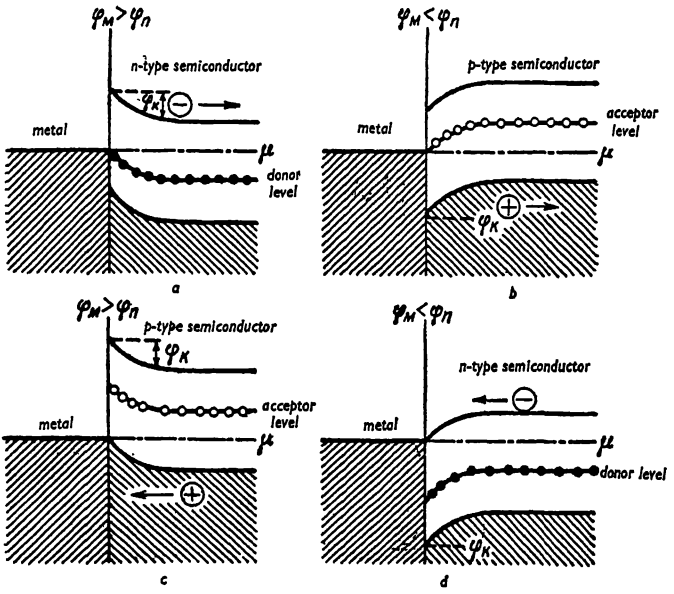


Fig. 81. Metal contact with n-type and p-type semiconductors. The formation of layers with increased (a, b) and reduced (c, d) electrical resistivity at the contact.

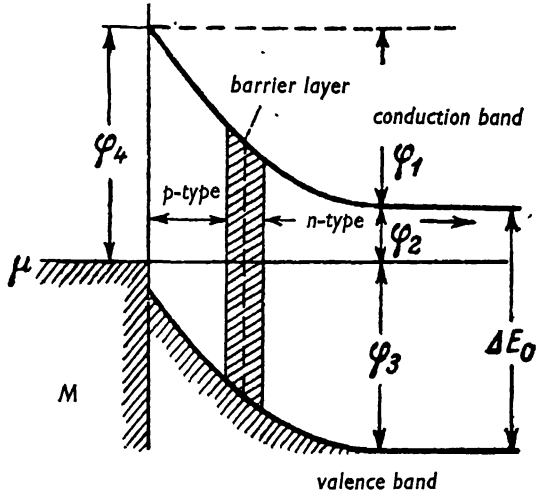


Fig. 82. Contact between a metal (M) and an n-type semiconductor in the case of a high contact potential difference. ΔE_0—width of the forbidden band; ϕ_1—the increase of potential energy of an electron near the surface of the semiconductor; ϕ_2—distance between the chemical potential level and the bottom of the conduction band; ϕ_3—distance between the chemical potential level and the top of the valence band; ϕ_4—work function at the semiconductor surface.

1. With large values of the contact potential difference V_0 the distance of the upper edge of the valence band from the chemical potential level in an n-type semiconductor may become less than the distance of the latter from the bottom of the conduction band. At all points of the surface layer where this situation arises the density of holes exceeds that of electrons and increases the electrical conductivity. At some distance away from the contact, however, a layer always arises in which the chemical potential level lies mid-way between the conduction and valence bands at a distance equal to half the forbidden band width (see Fig. 82). The electrical conductivity reaches in this layer its minimum value equal to that of the intrinsic semiconductor. On either side of this layer are regions of the opposite types of conductivity.

2. The second circumstance not allowed for in Davydov's theory is the increase of electrical conductivity in those regions of the contact layer in which the electric field exceeds the limit to which Ohm's law applies. The mean value of the electric field reaches 10^5 V/cm in a layer of 10^{-5} cm thickness at a potential difference of the order of 1 V.

Our investigations have shown that departures from proportionality between the current density and the electric field occur in the majority of semiconductors at fields of the order 10^4 V/cm.[23]

Of the two phenomena arising in the contact layers of semiconductors —the increase of resistivity in some cases and the increase of conductivity in others—the former manifests itself readily in the marked increase of resistance of the semiconductor placed between suitable electrodes. The second phenomenon, on the other hand, would make very little difference even with an infinitely high conductivity in a layer of 10^{-5} cm thickness. The effect of such a layer on the resistance of a specimen would be equivalent to the reduction of the specimen thickness by 10^{-5} cm.

The removal of mobile charges from the contact layer gives rise to an increase of its resistivity by several orders of magnitude, approaching the magnitude of the resistivity of the semiconductor when free from impurities (its intrinsic resistivity). Frequently, as in the case of cuprous oxide, the increase may be by as much as a factor of 10^6. The extra resistance due to a layer of 10^{-5} cm thickness is, therefore, equivalent to a layer of semiconductor of thickness $10^{-5} \times 10^6$ or 10 cm. If the actual thickness of a specimen is 1 mm the extra resistance at the electrodes amounts to 100 times the internal resistance. We have measured the resistance of a specimen, firstly, between gold or graphite electrodes having very high contact potentials and, secondly, between aluminium or magnesium electrodes, showing lower contact potentials. We observed in the second case a resistance 100 times larger than in the first.

The result is reversed if p-type cuprous oxide is replaced by n-type titanium dioxide. The resistance of a specimen of the latter material between gold electrodes is 100 times higher than between aluminium electrodes.

The relative increase in the electrical conductivity, caused by accumulation of free charges in the contact layer, does not attain such high values as the decrease due to sweeping away of free carriers. In fact, the composition of the contact field requires the introduction, into the layer, of approximately the same number of charges in both cases. In the case of removal of the free carriers from the layer they are replaced by immobile ions while the concentration of current carriers decreases by a factor of 10^6 or more. In the case of the accumulation of free carriers, on the other hand, their concentration increases only a few times.

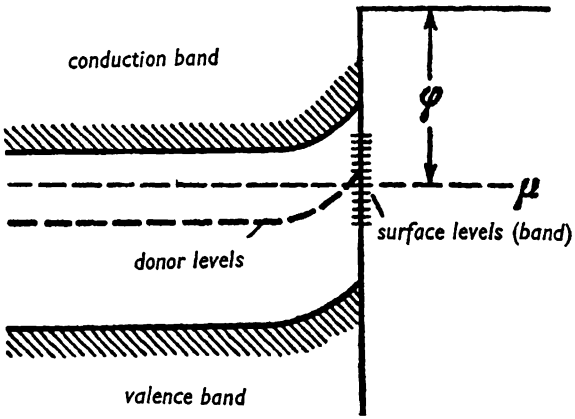


Fig. 83. The formation of a space charge layer in the vicinity of the surface of a semiconductor under the influence of surface charges.

Let us assume that the concentration of free carriers is n and that 10 n carriers per cc of the surface layers are required for the compensation of the field. In this case the electrical conductivity of the layer increases altogether ten-fold, whilst in the case of the opposite sign of the field almost all the mobile carriers are removed from the layer and are replaced by ions; the resistance of the layer may then increase thousands of times.

A layer of increased resistivity is called a barrier layer, one of increased conductivity is called an anti-barrier layer.

Both types of layer may arise in semiconductors, not only in contact with metals, but also at their boundaries in air. The necessary condition is the presence at the semiconductor surface of adsorbed ions or the appearance of charges filling surface energy levels of the crystal lattice.

Tamm has shown that, apart from the system of electric energy levels in the bulk of metal or semiconductor, the surface of the latter exhibits a second system of energy levels, different from the bulk system, which may fall in the forbidden band. These levels are always referred to as Tamm levels.

Surface charges give rise to a field in the surface layer which sweeps out, or draws in, mobile carriers in the same way as a contact field (Fig. 83). Both barrier and anti-barrier layers may arise according to the nature of the surface donor or acceptor levels, their position in the system of energy levels and the type of current carriers in the semiconductor.

Solid rectifiers

The barrier layers in semiconductors have properties which have resulted in practical applications: the resistance of a barrier layer is different for different directions of current flow. Application of an alternating voltage to the layer results in the flow of a higher current in one direction, the forward direction, than in the other, the reverse, direction.

The reasons for the appearance of such asymmetry may easily be seen. With one polarity of the applied potential the external field enhances the contact field, while with the opposite polarity it acts against the contact field. In the former case the thickness of the layer and its resistance increase; in the latter case the layer becomes thinner and, with a sufficiently high forward bias, its resistance becomes almost equal to that of the bulk semiconductor.

This explanation of the asymmetry of barrier layer resistance has been given by Davydov and Blokhnitsev[24] and has been developed in detail by Schottky.[25] It correctly predicts the direction of rectification and leads to a voltage–current relation both in the forward and reverse directions which is in qualitative agreement with the experimental data for rectifiers.

Frenkel' and I have studied the case of a potential barrier at a metal–semiconductor interface, the thickness of which is so small that it is partly transparent for electrons and holes.[26]

Consider what happens under these circumstances at the contact between a metal and an *n*-type semiconductor. In equilibrium, all the electrons present in the conduction band of the semiconductor may go over into the metal. The same number of electrons in the metal in levels above the Fermi level, which correspond to the conduction band of the semiconductor, may participate in an opposite current flow. Thus, the resulting current is equal to zero. Application of an external potential difference to cause electrons to flow from the metal into the semiconductor lowers the conduction band of the latter and the current from

the metal increases very strongly. Application of a potential difference of opposite polarity reduces the current from the metal. The flow from the semiconductor into the metal on the other hand, remains almost the same in both cases. However, this mechanism predicts a sense of rectification which is the same as is found in alternating current rectifiers, employing cuprous oxide, selenium, copper sulphide, i.e. *p*-type semiconductors. It is evident, further, that for the barrier to be transparent the layer thickness must not be larger than 10^{-7} cm, i.e. it may only arise in materials with a resistivity less than 0·1 ohm cm. In the case of germanium–metal contacts the thickness of the barrier layer often does not permit the tunnel effect mechanism. It is possible, however, that, apart from such cases, there may be others where the resistivity of the material does give rise to very thin barrier layers in which a tunnel effect may occur.

Van Geel attempted to explain the rectifying action of a barrier layer in terms of the effect of the high electric field, of the order 10^5 V/cm, appearing in the layer. In a medium of high dielectric constant fields of such magnitude might give rise to so-called electron emission from the surface of the metal. This explanation, like the previous one, leads to the wrong sign of rectification and is not verified by measurements of thickness of the barrier layer and of the electric field in it.

Zhuze evaporated a film of amorphous silica of variable thickness, followed by a metal film, on to the surface of cuprous oxide. He found that the most pronounced rectification (coefficient of rectification of the order 100) corresponded to a silica layer of about 10^{-5} cm thickness. Schottky suggested that this phenomenon could be explained on the assumption that the silica layer was porous and that the metal penetrating the pores acted as a large number of point contacts which were responsible for the rectification. It has been found, however, that rectification could be obtained even with thin layers of mica which one would not expect to be porous.[27]

We mention finally Bethe's theory of rectification at barrier layers with a thickness less than the free path length, in which the electrons traverse the space charge layer without collisions, as well as the theory of penetration of electrons through energy barriers of still smaller thickness.

Evidently none of these theories is capable of explaining the processes in technical selenium or cuprous oxide rectifiers which have a rectification coefficient (ratio of forward to reverse current) of the order of thousands or even tens of thousands. Such high values of the rectification coefficient were never encountered in metal–semiconductor contacts, which are usually expressed in terms of not many units. This has led to the hypothesis that the high coefficient of rectification arises, not at the interface between the semiconductor and the metal, but at the boundary

with another semiconductor having the opposite type of conductivity and a correspondingly high contact potential difference at the junction.

In cases where the *p*-type semiconductor has a higher contact potential than the *n*-type, as is commonly found in practice, barriers arise in the boundary layers of both. With current flowing in the forward direction, electrons and holes move towards the common boundary, recombine there and lead to the flow of a continuous current. With the reverse bias holes and electrons are swept from the boundary leaving behind a layer depleted of mobile carriers.

Here too, as in the case of metal–semiconductor contact, the barrier layer thickness increases with current flowing in the reverse direction and decreases with forward flow. Apart from this, the effect of the increasing conductivity with a high field acquires some importance. A. V. Ioffe, who has suggested this mechanism, carried out experiments with contacts between the semiconductors Cu_2O and TiO_2. She obtained values of rectification coefficients up to 20,000 while a simple contact between cuprous oxide and copper, or between selenium and the alloy used as electrode in selenium rectifiers, gave insignificant rectification, although the contact resistance was high and did show some dependence on the external potential. A good technical rectifier may only be obtained by means of complicated heat treatment in the case of cuprous oxide, or by passage of high reverse currents and simultaneous heating in the case of selenium rectifiers. Subsequent X-ray analysis proved that the heat treatment (forming) in the selenium rectifier gives rise to the formation of a layer of *n*-type cadmium selenide on the surface of the *p*-type selenium. The rectification in the presence of such a layer becomes independent of the contact potential of the electrode. This clearly shows that the rectification in this case takes place, not at the metal interface, but at the boundary between the *p*-type selenium and *n*-type cadmium selenide.

In the case of cuprous oxide, Lashkariev proved, by means of thermoelectric probing, the presence of an *n*-type oxide layer at the interface between the metal and the *p*-type oxide.

In germanium and silicon rectification takes place at the boundary between *p*- and *n*-type layers situated in the bulk of the same crystal.

We have seen in Fig. 82 that a strong curvature of the bands at the surface gives rise to a reversal of the conductivity type in the layer adjoining the metal. The transition from *p*- to *n*-type conduction occurs at a certain distance away from the contact and may lead to strong rectification, provided that it lies sufficiently deep to prevent a tunnel effect through the inversion layer adjoining the metal.

The first theoretical treatment of rectification at a boundary between *p*- and *n*-type semiconductors was given by Davydov. A more recent

detailed treatment by Gubanov deals with the case of thick diffusion layers.[28] Its results are in good agreement with experimental work by A. V. Ioffe. The case of thin layers, of the order of the free path length, has been dealt with by Bethe and Shockley.[29]

Davydov and Blokhnitsev have considered the rectification of current at a boundary between two semiconductors of the same type but with different electrical conductivities. Rectification is present even in this case too but it is quantitatively insignificant.

It is evident from the preceding exposition that the most effective arrangement for obtaining asymmetry of resistance consists of a p-type semiconductor in contact with an n-type semiconductor, provided that the contact potential of the former is higher than that of the latter.

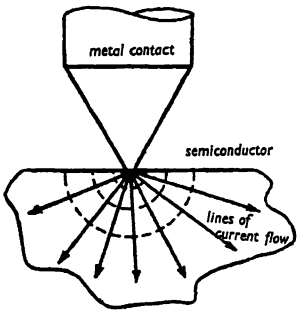


Fig. 84. Distribution of lines of current flow in the neighbourhood of a point contact.

Some peculiar phenomena occurring at the contact between p-type and n-type silicon carbide (including the emission of light on the passage of a current) were discovered and studied in detail by Lossev in the nineteen twenties, i.e. long before the appearance of modern theories of rectification.

Barrier layers commonly have thicknesses of the order 10^{-5} cm. Their presence readily manifests itself only when their resistivity exceeds the resistivity of the bulk material by many orders of magnitude. In particular, since the bulk of the conductor obeys Ohm's law and its resistance does not depend upon the direction of current flow, asymmetry may be introduced only by that part of the total resistance which is due to the barrier layer. The position changes, however, when the contact is restricted to the very small area of a metal point in contact with the semiconductor. Here, the chief contribution to the resistance is due to the spreading of the lines of current flow from the point contact into the bulk of the semiconductor (Fig. 84). It arises in a volume immediately adjoining the point contact and having dimensions comparable with the linear dimensions of the contact area, i.e. in a volume taken up by the

barrier layer. In this case the bulk of the semiconductor does not obscure the rectifying properties of the barrier layer and its properties are distinctly observable.

In fact, the first example of rectification of high-frequency currents was afforded by a metal point in contact with a semiconducting crystal—the so-called crystal detector of early radio amateurs.

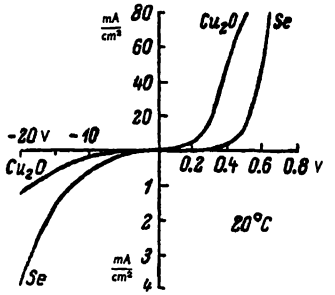


Fig. 85. Static voltage-current characteristics of cuprous oxide and selenium solid rectifiers.

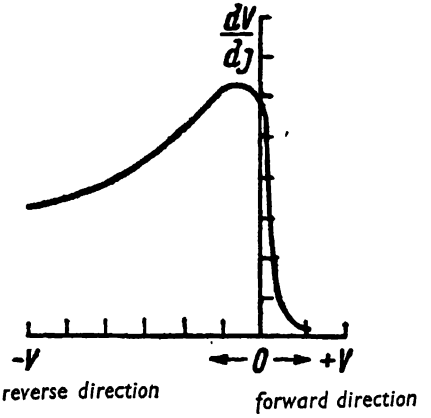


Fig. 86. The dependence of the incremental resistance of a semiconducting rectifier upon the applied voltage.

The voltage–current characteristics of typical cuprous oxide and selenium solid rectifiers are represented in Fig. 85. In a voltage range restricted to a few millivolts, the contacts obey Ohm's law and the resistance remains independent of the direction of current flow. Beyond this range, the current increases much more rapidly with voltage in the forward than in the reverse direction. Thus, the resistance of the barrier layer falls rapidly in the forward direction and changes only very slowly in the reverse, so that the coefficient of rectification increases

with the increasing voltage. At voltages in excess of a certain value the reverse resistance begins to fall and the rectification coefficient decreases accordingly. Fig. 86 shows the dependence of the incremental resistance dV/dj of the barrier layer upon the forward and reverse applied potential V.

The relation between resistance and thickness of the barrier layer, in the case of the diffusion barrier described above, differs from that for a layer of thickness L less than leV/kT (where l usually lies between 2×10^{-7} cm and 10^{-6} cm). An altogether different dependence upon thickness is obtained for semi-transparent energy barriers of a few Å thickness.

While the resistance of a diffusion barrier layer increases with its thickness L, the resistance of a layer which is traversed by electrons without collisions should be independent of thickness. In the latter case we find, assuming classical statistics, that the number dq of electrons passing through the barrier per second with normal components of velocities in the range v, $v+dv$ is given by:

$$dq = vn_0\left(\frac{m^*}{2\pi kT}\right)^{\frac{1}{2}}\exp\left(-\frac{m^*v^2}{2kT}\right)dv. \tag{138}$$

Here n_0 is the total density of electrons in the conduction band.

The surface layer represents an energy barrier which is non-transparent for electrons, and has the height $e(V_C \pm V)$, where V_C is the contact potential difference and V is the externally applied potential difference. Only the electrons with kinetic energy $\frac{1}{2}m^*v^2$ greater than $e(V_C \pm V)$ are capable of surmounting the barrier. Defining the limiting velocity v_0 by the condition:

$$\tfrac{1}{2}m^*v_0^2 = e(V_C \pm V),$$

the number of electrons traversing the barrier is:

$$q = \int_{v_0}^{\infty} dq = n_0\left(\frac{kT}{2\pi m^*}\right)^{\frac{1}{2}}\exp\left(-\frac{e(V_C \pm V)}{kT}\right). \tag{139}$$

No current should flow in the system for V equal to zero. Consequently, the current of electrons q must be compensated by an equal current q_0 given by:

$$q_0 = n_0\left(\frac{kT}{2\pi m^*}\right)^{\frac{1}{2}}\exp\left(-\frac{eV_C}{kT}\right). \tag{140}$$

This current remains constant even for $V \neq 0$ and the resulting current J is:

$$J = e(q-q_0) = \frac{en_0}{\sqrt{(4\pi)}}\sqrt{\left(\frac{2kT}{m^*}\right)}\exp\left(-\frac{eV_C}{kT}\right)\left[\exp\left(\pm\frac{eV}{kT}\right)-1\right]. \tag{141}$$

With $V_C = 0$, the current is

$$J = \frac{en_0}{\sqrt{(4\pi)}}\bar{v}. \tag{142}$$

This corresponds to the transit of all the n_0 valence electrons across the barrier layer with a mean velocity

$$v_1 = \frac{\bar{v}}{\sqrt{(4\pi)}} \tag{143}$$

where $\bar{v}$ is the root mean square of the thermal velocity of the conduction electrons. For increasing negative values of V the current decreases rapidly.

Equation (141) gives a good approximation to the experimental results for germanium rectifiers in which the thickness L is about 5×10^{-6} cm for a resistivity 0·1 ohm cm.

In the case of a diffusion barrier layer with complete ionization of impurities we get:

$$J = en_0 u\mathscr{E} \exp\left(-\frac{V_C}{kT}\right)\left[\exp\left(\pm\frac{eV}{kT}\right) - 1\right]. \tag{144}$$

Here u is the electron mobility and $\mathscr{E}$ equal to $(V_C \pm V)/L$ is the mean value of electric field strength in the layer. $\mathscr{E}$ depends both on the applied potential V and the thickness L of the layer.

The theory of rectifiers concludes that the thickness of a barrier layer increases with reverse bias and decreases with forward bias. This conclusion is confirmed directly by experiment. In view of the fact that the resistivity of the barrier layer is very much higher than that of the bulk semiconductor, the former appears as the dielectric between two condenser plates consisting of the metal electrode and the semiconductor. The measurement of the capacitance of this condenser gives the thickness of the dielectric immediately. Furthermore, it is possible to determine the resistivity of the barrier layer, as a function of distance from the electrode, by varying the frequency of the measuring current.

The theory of diffusion equilibrium predicts that the thickness of the barrier layer is proportional to the square root of the resistivity of the bulk semiconductor.

In semiconductors having a resistivity less than 10^{-2} ohm cm and a mobility of the order 100 cm²/V sec, the barrier layer thickness is less than 10^{-7} cm. Electrons can traverse this barrier by the tunnel effect and the barrier resistance is insignificant. In the case of resistivities of the order 10^{-2} to 10^{-1} ohm cm the barrier thickness exceeds 10^{-7} cm but frequently is less than the mean free path of electrons. In these cases electrons traverse the barrier layer without collisions but the forward and reverse currents are different. In semiconductors with resistivities between 10 and 100 ohm cm the barrier layer thickness exceeds 10^{-6} cm.

Here the electronic distribution within the barrier is determined by the equilibrium between the electric field at the contact and the diffusion due to a concentration gradient. Finally, in insulators with resistivities in excess of 10^9 ohm cm the barrier layer thickness is measured in millimetres or centimetres, the internal electric fields are weak and the concentration gradients low. Here the barrier layer gradually loses its significance; frequently it extends over the entire volume of the insulator. Equilibrium with a metal electrode is reached by altering the potential of the whole insulator and this is an exceedingly slow process. These phenomena have not yet been studied experimentally.

Slow time-dependent phenomena in barrier layers

Experiments on barrier layers show that the forward resistance decreases, and the reverse resistance increases, with time, leading to a gradual increase of the rectification coefficient. In the absence of an externally applied electromotive force, the contact resistance increases with the time after the formation of the contact (Fig. 87).

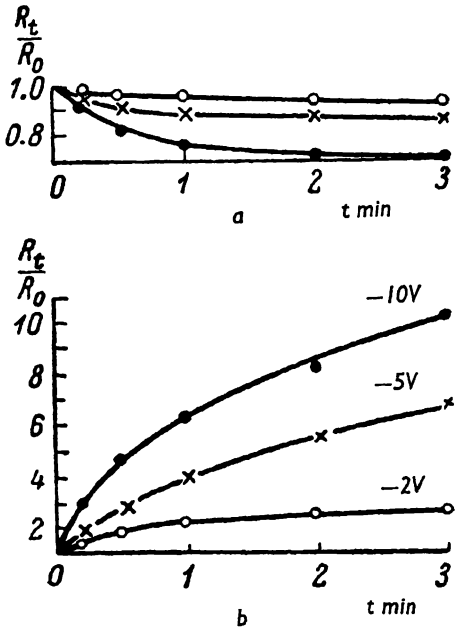


Fig. 87. Variation with time t of the barrier layer resistance of a Cu_2O-TiO_2 rectifier in the forward direction (a) and reverse direction (b).

After a detailed study of the phenomenon, A. V. Ioffe and I[30] have ascribed it to the migration of ions forming the space charge of the barrier layer. The mobility of ions in the bulk semiconductor is less than

that of electrons by a factor 10^8–10^{10}. Thus, although the density of ions is equal to the density of electrons or holes (and may be considerably higher in the case of compensated impurities), the former do not contribute appreciably to the electric current.

In the barrier layer, however, the conditions are especially favourable for the appearance of ionic conductivity. Firstly, the concentration of ions is higher there, since the impurities in the barrier layer are almost fully ionized, while in the bulk semiconductor the ionization is only partial. Secondly, the electrical field acting on the ions attains values of the order 10^5 V/cm. Thirdly, the paths of migration within the barrier layer are very small—of the order 10^{-5}–10^{-4} cm; it is sufficient for an ion to move through this distance in order to emerge from the barrier layer. The concentration of ions in the barrier layer of semiconductors, such as e.g. cuprous oxide, is of the order 10^{20} cm^{-3}. With a mobility of 10^{-10} cm^2/V sec and a layer thickness 10^{-4} cm, 10^{14} ions leave the layer per second, while their total number in the barrier layer is 10^{16}.

Using the simple model described above, it was possible to analyse the variation with time of the concentration of ions in the barrier layer, and the accompanying variation of barrier layer resistance. It was found, in full agreement with experiment, that the resistance increases in the reverse direction and decreases in the forward direction. The time-dependence of these changes also agrees well with theory. The fact that the resistance increases with one sign of the internal field and decreases with the opposite sign suggests that this phenomenon is due to the motion of charges and not of neutral atoms. On the other hand, the slow nature of the process, lasting for minutes and hours, indicates that it is due to ions and not to electrons.

The electrical character of the process is finally stressed by the appearance, in the barrier layer, of a polarization space charge after prolonged passage of a reverse current. The magnitude of this space charge is also in agreement with the results of the theoretical analysis.

The mobility of ions may be estimated from the measurements of the time-dependence of this process, on the assumption that the latter is due to the migration of ions. The value of their mobility in titanium dioxide was found to be 10^{-10} cm^2/V sec. A similar order of magnitude was obtained for the mobility of barium ions in BaO using radioactive tracer techniques. It follows, therefore, that in electronic semiconductors the mobilities of ions are negligible in comparison with the mobilities of holes and electrons, which lie between 0·1 and 10,000 cm^2/V sec. The exceedingly small mobilities may, nevertheless, be determined in the barrier layers.

The previously described phenomenon of forming of selenium rectifiers is, most likely, closely associated with ionic migration in the

barrier layer. The process of ionic migration is completed when not only the drift of electrons and holes in the electric field is compensated by diffusion, but also when the same compensation is achieved for ions in the barrier layer. Furthermore, it is likely that barrier layer formation is not completed even on attainment of equilibrium for electrons, holes and ions. Large densities of neutral impurity atoms are present in the bulk semiconductor, while in the barrier layer they are transformed into ions and are frequently removed from it by the electric field at the contact. The process of equalization of impurity concentrations between the bulk semiconductor and the barrier layer is much slower than the drift of ions in the barrier layer. It is possible that diffusion of impurities is responsible for the so-called ageing of rectifiers for periods extending over months and years. This process has not yet been studied either experimentally or theoretically.

p–n junctions

A major advance in the technology of rectification has been made possible through the replacement of contacts between *p*-type and *n*-type crystals by *p–n* junctions. A *p–n* junction consists of an abrupt transition from acceptor to donor impurity within the same continuous crystal lattice. The "contact" between the *p*-type and *n*-type regions is made without the intervention of crystal surfaces, thus ensuring freedom from the attendant imperfections and irregularities. The properties of the material on both sides of the *p–n* junction may be very accurately controlled and materials with very high mobilities have become available (silicon with 1200 cm^2/V sec and germanium with 3800 cm^2/V sec have replaced selenium with 1 cm^2/V sec and cuprous oxide with 50 cm^2/V sec).

As a result, the frequency range, the current-carrying capacity and the efficiency of rectifiers have been increased enormously. Furthermore, entirely new possibilities have been opened through the application of *p–n* junctions to amplification and the generation of high frequency signals.

Various methods of *p-n* junction preparation have been developed. One method takes advantage of the possibility of introducing impurities from the melt in the process of crystallization. An impurity is first added to the melt during the growth of a single crystal and the addition of another impurity follows after some time. When these impurities enter the single crystal they make it respectively *p*- or *n*-type. Frequently both impurities are added simultaneously and the preponderance of one or the other in the solid phase is achieved by the variation of the growth rate of the crystal.

Alloying represents the method most frequently employed industrially. Thus, for example, the alloying of an indium pellet on to

n-type germanium results in the formation of a p–n junction. A germanium-indium alloy is formed on heating to some 550° C. Subsequent cooling produces an abrupt, almost step-function type, transition from n- to p-type material at the former solid–liquid interface.

Impurities may be introduced by diffusion from the surface. In the presence of an initial concentration N_0 of impurities per unit area of the surface layer, the concentration in the bulk, after time t, at a distance x from the surface, is given by:

$$N = \frac{N_0}{\sqrt{(\pi Dt)}} \exp\left(-\frac{x^2}{4Dt}\right). \tag{145}$$

The diffusion coefficient D depends upon temperature according to the formula:

$$D = D_0 \exp(-E/kT). \tag{146}$$

Diffusion in germanium is characterized by the following values of the parameters D_0 and E:

TABLE 6

	D_0 (cm²/sec)	E (eV)	
As (donor)	7.10^{-1}	2·2	
In (acceptor)	4.10^{-3}	2·2	
Sb (donor)	7.10^{-1}	2·2	according to different measurements
	10	2·5	
Zn (acceptor)	9.10^{-2}	2·2	
	6.10^{-1}	2·5	
Thermal defects and Cu	2.10^{-2}	0·53	

A particularly convenient method of obtaining p–n junctions of desired characteristics consists in the diffusion of donor and acceptor impurities from opposite surfaces of a thin germanium wafer. This method maintains the transition region almost free from impurities, and enables the formation of so-called p–i–n structures.

It is possible to obtain a p–n junction on the surface of single crystal n-type material by bombardment with ions, forming acceptor impurities, or with fast electrons which also change the conductivity type.

The conditions of recombination of free carriers, both in the bulk semiconductor and its surfaces, are of special importance. The direct recombination of an electron with a hole has a low probability. An amount of energy equal to the width of the forbidden gap is liberated at one stage in the process. The lattice, however, can only absorb thermal energy in the form of phonons with energy $h\nu$ where ν is the frequency of a normal vibration mode of the lattice. The magnitude of ν does not exceed 10^{13} sec^{-1} and, thus, the energy of a phonon does not exceed

$6{\cdot}6 \times 10^{-14}$ erg or 0·04 eV. Usually this energy is much less than 10^{-1} eV.

In the case of a forbidden gap width of approximately 1 eV, the recombination of a hole–electron pair would require the simultaneous absorption by the lattice of some tens of phonons. Quantum theory shows that the probability of a multi-quantum transition decreases with increasing number of simultaneously interacting quanta.

This explains why the most important mechanism of recombination consists in the trapping of one type of carrier at a lattice defect and the subsequent attraction of the localized charge of the opposite type of carrier. This mechanism of recombination does not involve multi-quantum transitions. The momentum and energy of the carrier are transmitted to the defect which, having a large mass, is capable of absorbing a considerable amount of energy by increasing its local temperature. The temperature is restored by thermal conduction, and the energy absorbed by the defect is, thus, dissipated in the lattice.

In the event of an increase of the carrier densities above the thermal equilibrium values, whether by illumination or by other means, the probability of recombination is determined by the lifetime τ_0 of the excess carriers. τ_0 depends upon the nature and concentration of the defects and may reach values as high as 1000 μsec in high-purity germanium. In view of the fact that the thermal velocity of a free carrier is of the order 10^7 cm/sec, it follows that, during the time of its free existence, a carrier traverses a distance of 10^4 cm. With a free path length of 10^{-5} cm, the carrier may be scattered 10^9 times before it recombines. Since the diffusion constant of electrons is of the order 100 cm^2/sec, the diffusion length L equal to $\sqrt{(D\tau_0)}$ in a germanium specimen reaches 3 mm. It should be understood that all the magnitudes quoted above have a statistical character and the numerical values are only indicative of the order of magnitude at room temperature.

The highest concentration of defects, both structural as well as chemical, is to be found at the surface. The probability of recombination at the surface is accordingly very high, even to the point that almost every carrier reaching the surface may end its free existence there. The probability of surface recombination depends to a large extent upon the surface treatment. It is very high for ground surfaces and decreases considerably upon etching in suitable reagents which dissolve the crystal.

The treatment for low surface recombination acquires particular importance in the region of a p–n junction near the surface of the crystal. Recombination at the surfaces reduces τ_0 and, hence, also the diffusion length L which, as we shall see, determines the quality of devices based on p–n junctions.

We shall now describe the properties of p–n junctions. We have to take into account the existence, in an n-type semiconductor, of a certain density of holes as well as free electrons. Let us denote the density of holes here by n_+ and the density of electrons by n_-. Under equilibrium conditions

$$n_- n_+ = n_0^2$$

where n_0 is the concentration of free electrons or holes in an intrinsic sample at a given temperature T. We shall use capital letters P and N to denote the conductivity type, so that n_{+N} denotes the density of holes in an n-type semiconductor and n_{+P} the hole density in a p-type semiconductor. The term majority carriers denotes electrons in n-type, and holes in p-type, semiconductors. Electrons in p-type material and holes in n-type material, on the other hand, are referred to as minority carriers.

The equilibrium of electrons and holes between conductors in contact, or between different parts of the same semiconductor, requires the equality of their chemical potentials μ. In semiconductors with a high dielectric constant, as, for example, silicon and germanium, the donor levels fall at a distance of only a few hundredths of an eV from the conduction band edge. The acceptor levels are at a similar distance from the valence band. At ordinary temperatures for which kT is about 0·03 eV all impurities are fully ionized. In an n-type semiconductor the chemical potential falls above the middle of the forbidden gap, close to the conduction band. In a p-type semiconductor the chemical potential level lies below the centre of the forbidden gap near the valence band.

The equalization of the chemical potential levels at the boundary between n- and p-type regions is brought about by a transfer of electrons into the p-type region and of holes into the n-type region. The impurity ions left behind in the transition region form a negative space charge in the p-type part of the semiconductor and a positive space charge in the n-type layer. A potential difference of a magnitude sufficient to prevent the transfer of majority carriers from one part into the other is, therefore, set up between the p- and n-type regions.

Diffusion of majority carriers tends to take place across the transition region owing to the fact that $n_{-N} \gg n_{-P}$ and $n_{+P} \gg n_{+N}$. Assuming that the concentration is a function of the x-coordinate only, we have at any transverse plane:

$$j_N = D_N e \frac{dn_-}{dx},$$

$$j_P = -D_P e \frac{dn_+}{dx}.$$

After the establishment of equilibrium this diffusion is counterbalanced by a drift current of charges due to the electric field $\mathscr{E}$:

$$D_N \frac{dn_-}{dx} = -n_- u_N \mathscr{E},$$

$$D_P \frac{dn_+}{dx} = n_+ u_P \mathscr{E}.$$

The same electric field $\mathscr{E}$ compensates for both the flow of electrons from n-type and the flow of holes from p-type regions:

$$\frac{D_N}{u_N} \cdot \frac{1}{n_-} \cdot \frac{dn_-}{dx} = -\frac{D_P}{u_P} \cdot \frac{1}{n_+} \cdot \frac{dn_+}{dx}.$$

But

$$\frac{D}{u} = \frac{kT}{e};$$

hence

$$\frac{dn_-}{n_-} = -\frac{dn_+}{n_+} = -\frac{e}{kT} \mathscr{E} dx.$$

Integration over the thickness of the transition region gives:

$$\frac{n_{-N}}{n_{-P}} = \frac{n_{+P}}{n_{+N}} = \exp\left(\frac{eV}{kT}\right) \tag{147}$$

where V is the potential difference between the p- and n-layers.

In the boundary layer itself, n_+ and n_- are determined only by the intrinsic conductivity of the material,

$$n_- = n_+ = n_0.$$

Fig. 88 shows the energy level diagram of a p–n junction in germanium.

On the application of an external potential difference V_{ex} to such a system, most of the potential drop is taken up by the p–n junction, since its resistance is very much higher than the resistance of the remaining parts of the crystal. A flow of current is caused by the disturbance of equilibrium between the p- and n-regions.

It would be entirely wrong, however, to determine the magnitude of the current from the magnitude of the resistance of the transition region. This procedure would only be justified in the case of a transition region of such thickness that the concentrations of free carriers in any section would be established independently of the properties of the neighbouring sections. The resistances of a series of mutually independent sections could then be summed to give the overall resistance of the whole region. Different conditions obtain, however, in the transition region of a germanium p–n junction.

The barrier layer thickness d is small in comparison with the free path length l of the electrons, which may reach the magnitude of

1 micron, while the diffusion length L is thousands of times larger, measuring up to 1 mm. Therefore, the properties of any particular section of the transition region do not depend solely upon the temperature and the concentrations of donors and acceptors in this section, but also, to a large extent, upon the carrier concentrations in the surrounding parts of the material, up to a distance L away from the section in question. We may go to the extent of stating that any particular section serves only as a transit plane for the drift and diffusion currents which traverse it under the action of the electric field and the concentration gradients.

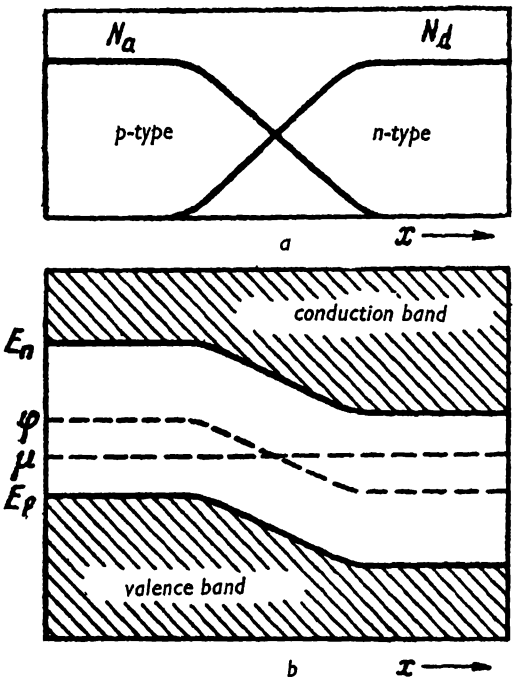


Fig. 88. (a) Distribution of impurities in a p–n junction; (b) the corresponding distribution of electrostatic potential with zero applied bias.

If the external field is directed in the same sense as the contact field, the probability of transfer of majority carriers—electrons from n-type, and holes from p-type, regions—becomes even smaller. The current in this case is due to minority carriers—electrons from the p-region and holes from the n-region—which move under the combined influence of both the contact field and the externally applied field.

The density of minority carriers is the higher, the lower the density of majority carriers, and vice versa. The current in the reverse direction,

is, thus, the higher, the larger the concentration of minority carriers and consequently, the lower the concentration of majority carriers.

At not excessively high values of the reverse bias V, which, however, satisfy the condition $V \gg kT/2$, the reverse current depends only very little on the external potential difference. This is because of the fact that the current is due primarily to the flow of minority carriers which are equally able to cross the transition region of the p–n junction, whatever the magnitude of the external field assisting their flow. The current may rise slightly with increasing bias owing to a number of causes. Among these are the increase of junction temperature, the widening of the space charge region with increasing voltage, the onset of impact ionization or the tunnel effect, and the fact that the electric field frequently extends beyond the transition region and increases the range from which carriers are collected.

The passage of a current alters the equilibrium concentrations of holes and electrons. Free carriers present within the range of a diffusion length L equal to $(D\tau_0)^{\frac{1}{2}}$ diffuse and reach the p–n junction. L has different values for holes and electrons but, in germanium, both are of the same order of magnitude. It should be understood that diffusion is a statistical process and, rather than thinking in terms of a sharply defined diffusion length L, it is more correct to use the probability $\exp(-x/L)$ of a carrier traversing, by diffusion, a distance x. Of all the carriers present within a distance L, a proportion $\exp(-1)$ or 0·37 will reach a given point by diffusion.

The concentration n of free electrons in a layer in the p-type region, at a distance x from the p–n junction, may be expressed as:

$$n_- = n_{-P} + n_{-1} \exp\left(-\frac{x}{L_N}\right) \tag{148}$$

where

$$n_{-1} = n_{-P}\left[\exp\left(\frac{eV_{ex}}{kT}\right) - 1\right].$$

Similar relations apply to the dependence of n_+ on n_{+N}.†

In the case of a large positive external potential difference V_{ex}, the concentration n_{-1} is large. For large negative values of V_{ex}, on the other hand, n_- is almost zero since n_{-1} is nearly equal to n_{-P}.

The electronic component of the current J_N traversing the p–n junction is given by:

$$J_N = eD_N\left(\frac{dn_-}{dx}\right)_{x=0}. \tag{149}$$

† W. Shockley, *Electrons and Holes in Semiconductors* (Toronto, New York, London, 1956), p. 309.

Substituting for n_- from equation (148) and writing D_N equal to L^2/τ_N, we obtain

$$J_N = \frac{eL_N n_{-P}}{\tau_N}\left[\exp\left(\frac{eV_{ex}}{kT}\right)-1\right],$$

where n_{-P} is the minority carrier concentration at a large distance away from the junction. The total current across the junction is composed of electron and hole currents flowing in opposite directions

$$J = \left(\frac{eL_P n_{+N}}{\tau_P}+\frac{eL_N n_{-P}}{\tau_N}\right)\left[\exp\left(\frac{eV_{ex}}{kT}\right)-1\right]. \tag{150}$$

For large negative values of V_{ex} the value of the second bracket tends to -1 and the reverse current may be regarded as being due to the flow across the junction of all the minority carriers generated within a diffusion length L away from the junction.

For small values of $\pm eV_{ex}/kT$ the square bracket term approximates to $\pm eV_{ex}/kT$ and the current density may be expressed by:

$$J = \frac{e^2}{kT}\left(\frac{L_P n_{+N}}{\tau_P}+\frac{L_N n_{-N}}{\tau_N}\right)V_{ex}. \tag{151}$$

But $$\frac{e^2}{kT}\cdot\frac{L_N n_{-P}}{\tau_N} = \frac{u_N e n_{-P} L_N}{D_N \tau_N} = \frac{e u_N n_{-P} L_N}{L_N^2} = \frac{\sigma_-}{L_N}.$$

Hence, for small values of V_{ex},

$$J = \left(\frac{\sigma_+}{L_P}+\frac{\sigma_-}{L_N}\right)V_{ex} \tag{152}$$

where σ_+ and σ_- are the conductivities due to minority holes and electrons in the n- and p-type regions respectively.

For large positive values of eV_{ex}/kT the current increases exponentially. Since eV_{ex}/kT is nearly equal to 1 for V_{ex} as low as 0·026 V, the forward current begins to increase rapidly for potential differences in excess of a few hundredths of a volt.

Fig. 89 shows the voltage–current characteristic of a germanium point-contact diode and Fig. 90 that of a junction rectifier.

In the reverse direction the current initially remains practically constant and independent of V_{ex} (more accurately, it increases very slightly with voltage). At very high values of V_{ex}, a rapid increase of current sets in, leading to breakdown of the junction.

The increase of current with voltage is frequently followed by a further increase, accompanied by a lowering of potential difference across the junction region. The differential resistance dV/dJ assumes a negative value (Fig. 91). Before the appearance of vacuum valves, negative voltage–current characteristics were employed for the generation of oscillations in the "singing arc".

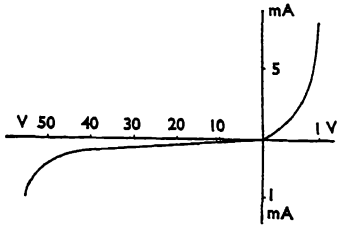


Fig. 89. Characteristic of a point-contact diode.

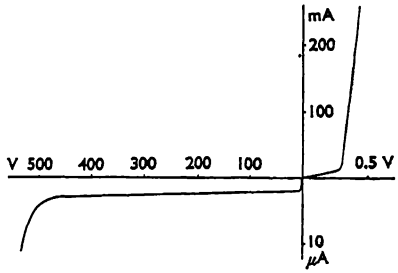


Fig. 90. Voltage-current characteristic of an alloyed germanium p–n junction.

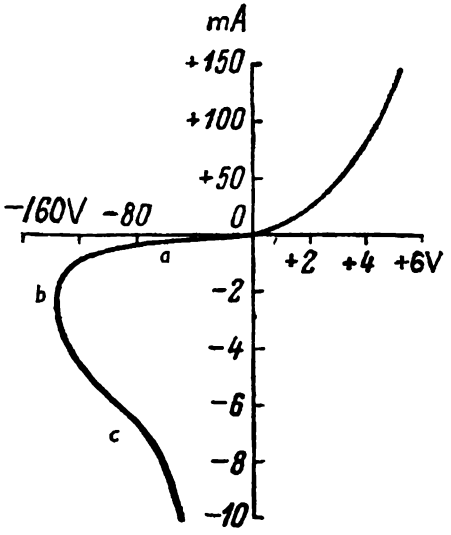


Fig. 91. Voltage-current characteristic of a metal contact on n-type germanium: a) low-voltage part of the reverse branch of the characteristic; (b) transition part; c) negative resistance region.

Equation (150) accounts well for the forward characteristic at not very large values of V_{ex}, i.e. at low current densities. A more detailed analysis of the problem takes into account not only the conditions at the p–n junction itself, but also the effect of the electric field on the motion of carriers outside the junction region as well as the part of the potential-drop within the junction.†

Rectifiers and transistors

The crystal detectors of early wireless enthusiasts (which, in fact, were crystals of lead sulphide, zincite, silicon carbide or pyrites in contact with a lightly pressed metal wire) have again found widespread applications in radio technology as germanium and silicon diodes. Unlike the unstable early crystal devices, in which a sensitive spot had to be found afresh each time, the germanium diodes are distinguished by stable operation.

We have seen that the boundary between n- and p-type semiconductors has good rectifying properties, provided that the contact potential of the p-type material is higher than that of the n-type region. The latter condition is automatically satisfied if the two semiconductors in contact are, in fact, prepared from the same material but with different impurities.

It may be assumed that the search for a sensitive spot in the early crystal devices amounted to the location of an area in which the surface layer had the opposite type of conductivity from the bulk of the material. In germanium point contact diodes such layers are formed artificially by suitable treatment and forming—passage of current across the contact. It is regarded as certain that the considerable heating which arises under these conditions, together with the high electric fields due to the forming pulse, lead to the formation, on n-type germanium, of a localized p-type region surrounding the metal point contact. Similar processing leads to the formation of rectifying contacts on p-type germanium and silicon. Apart from special etching of the surface, the technique either uses electrodes with suitable impurities or deposits these impurities on the surface of the semiconductor. During the forming pulse the impurities penetrate into the material and change its type of conductivity. Thus, for example, an admixture of Sb is used for p-type silicon and Al for n-type silicon, in agreement with the properties of these impurities as described previously.

The junction between p- and n-type germanium or silicon may, thus, fulfil all the requirements for a good rectifier.

† Report of the Meeting on Semiconductors held by the Physical Society in collaboration with British Thomson-Houston Ltd., Rugby, in April 1956.

Germanium point contact diodes are distinguished by good stable characteristics, low cost, small dimensions, negligible capacitance and rapid transit of carriers across the barrier. For these reasons they have found wide application in centimetric wavelength radar systems and in digital computers where tens of thousands of diodes must be accommodated in a small space.

Further advantages consist in mechanical rigidity, long life and the absence of heating circuits. Against these must be set the difficulties of production, due to the present unsatisfactory standardization of technique, and the scatter of device parameters which leads to the necessity for grading of the finished articles.

Another disadvantage of crystal detectors consists in the strong temperature dependence of their characteristics. Thus, in practice, germanium diodes can hardly be used at temperatures above 60° C. Silicon diodes may operate above 100° C and selected specimens even up to 200° C, this being due to the lower intrinsic conductivity of silicon. This compels us to search for new materials, with higher forbidden energy gaps, for the development of rectifiers suitable for high temperature work. One such material may be SiC with an energy gap of approximately 3·1 eV.

Semiconductor devices suffer from the disadvantage of a relatively higher noise level than vacuum valves.

The voltage–current characteristics of germanium and silicon point contact rectifiers are somewhat different: in silicon rectifiers strong asymmetry sets in even at very low voltages, while in germanium a bias of 0·1–0·2 V is necessary for good rectification.

Germanium rectifiers exhibit one more phenomenon: we have seen already that at certain reverse voltages a further increase of current leads to a reduction of voltage due to increased heating. The differential resistance R_d equal to dV/dJ becomes negative.

The phenomenon of falling characteristic was discovered by Losev as early as 1922 for the case of steel wire contacts on crystals of zincite and some other semiconductors. He called this device a "crystadyne". This phenomenon did not vanish at high frequencies so that a crystadyne has all the properties of a regenerative valve receiver and low power generator. Crystadyne receivers were widely used by radio amateurs.

Apart from germanium and silicon point contact diodes, large area junction rectifiers are also currently in production. The p–n junction in this case extends over an appreciable area. Large area diodes allow, essentially, the rectification of much higher currents than point contact types. The current density and the reverse voltage are also much higher in these junction diodes than in selenium and cuprous oxide rectifiers. Silicon rectifiers allow the higher reverse voltages to be applied. The

frequency response of junction rectifiers is limited by their capacitance and is much lower than that of point contact rectifiers.

It is evident from this very brief review of germanium and silicon rectifiers that, even at the present time, they possess a number of advantages over vacuum devices and have specific fields of application. Germanium and silicon diodes have characteristics which are undoubtedly superior to the old crystal detectors of the early days of radio communication. This advance has been made possible by deliberate introduction of desirable electrical properties through the suitable distribution of p–n junctions.

In connection with p–n junctions we should acknowledge the priority of Losev, who in the latter years of his life (1931–39) investigated visible layers on silicon carbide crystals having opposite types of conductivity. By passing current through the barrier layer so formed Losev observed the emission of visible light and was able to establish the correct quantum relation between the frequency ν of the emitted light and the energy gained by an electron in its passage across the barrier potential difference V:

$$h\nu = eV.$$

The visible layer, consisting of quartz or n-type SiC, was too thick for an electron to penetrate it without collision and, thus, acquire the necessary energy for light emission. The thickness of the barrier layer, however, formed at the interface between the transparent layer and the dark silicon carbide, i.e. at the boundary between the p- and n-type silicon carbide, was probably not larger than one free path.

In this way Losev not only observed rectification at a p–n junction in silicon carbide, but also discovered and correctly explained the emission of light accompanying the passage of current across the junction. Losev's discovery has recently been confirmed in the U.S.A.

Apart from the rectifying effects in germanium and silicon diodes, which correspond to diode valves, germanium devices are also used as radio amplifiers (transistors) replacing vacuum triodes.

Semiconductor triodes (transistors) are made, both with point contacts, using metal whiskers, and as junction devices with appreciable junction areas. The principle of operation of a semiconductor triode may be best understood with reference to the junction type.

Fig. 92 shows diagrammatically the structure of a transistor and also one of the possible circuit connections.

Unlike a diode, a transistor has two p–n junctions, usually at a separation small compared with the diffusion length. Both p–n–p and n–p–n transistors are in existence. One of the junctions, called the emitter, is biased in the forward direction; the other is biased in the

reverse direction and is referred to as the collector. The region between the two junctions is called the base.

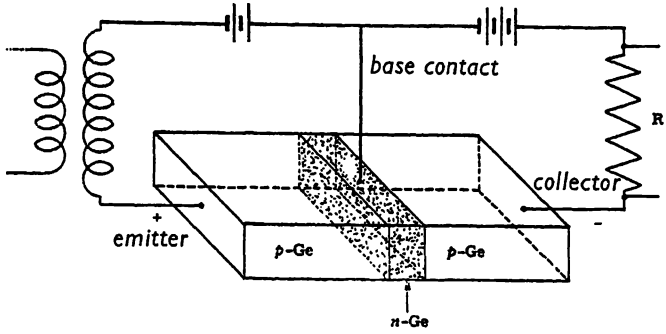


Fig. 92. The structure and circuit connections of a p–n–p junction transistor.

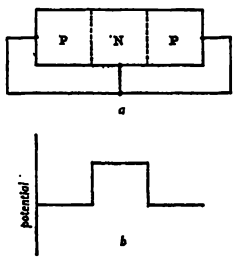


Fig. 93. (a) The diagram of a junction transistor in the absence of external sources of bias; (b) potential distribution.

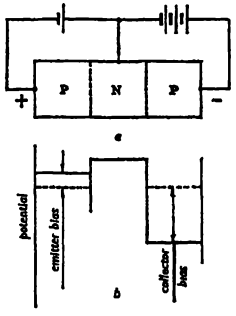


Fig. 94. (a) Diagram of a p–n–p junction transistor with the arrangement of external biases; (b) potential distribution corresponding to those conditions.

In the absence of external sources of bias, contact potentials or diffusion potentials appear across the p–n junctions, as shown in Fig. 93.

The external biases are applied, as shown in Fig. 92, in the sense of lowering the emitter–base barrier height and increasing the collector–base barrier height. The corresponding potential distribution is shown diagrammatically in Fig. 94.

The transport of current from the emitter into the base occurs by means of holes flowing from the p-type region into the n-type base and electrons flowing in the opposite sense. The resistivity of the p-type emitter region is made very much less than that of the n-type base, so that the concentration of holes in the emitter is vastly higher than the concentration of electrons in the base. The current from emitter to base

is, therefore, carried almost exclusively by holes. Under normal conditions the externally applied biases are impressed wholly across the respective *p–n* junctions, so that the electric field in the base region is negligibly small. The holes injected into the base from the emitter are, therefore, transported further mainly by diffusion. As shown in Fig. 94, the electric field at the collector junction acts in such a way as to remove into the collector any holes arriving at that junction. Thus, the concentration of holes at this junction may be regarded as negligible in comparison with their concentration at the emitter junction. In view of the fact that the base width is much less than the diffusion length for holes, almost the entire hole current, entering the base from the emitter, arrives at the collector. For this reason any fluctuations of current in the emitter circuit give rise to similar fluctuations of current in the collector circuit. The emitter fluctuations may be due to any cause whatever. For example, as shown in Fig. 92, they may be induced in the secondary winding of a transformer connected in the emitter circuit. With this mode of operation the ratio of the currents in the two circuits is very close to unity. However, the emitter junction operates in the forward direction and a given change of current may require only a small change of voltage, as may be inferred from the characteristic of Fig. 90. The same change of current in the collector branch, on the other hand, may give a much larger change of voltage over the load resistance R, as shown in Fig. 92.

The differential collector resistance is very high since the collector junction is operated in the reverse direction. It is possible, therefore, to use a high load resistance R in the collector circuit and to obtain large voltage fluctuations over this resistance as a result of small fluctuations of voltage in the emitter circuit. The battery included in the collector circuit provides the source of energy for the amplification of the signal which is induced in the emitter circuit. Other modes of transistor operation enable the amplification of current as well as voltage.

The power handling capacity of devices using large area junctions is substantially higher than that of point contact devices, but their frequency range is more restricted.

Present-day junction devices (*p–n–p* or *n–p–n*) have a power dissipation of up to 2 W and, most recently, units with output powers of up to 20 W have been announced. Furthermore, it should be pointed out, that junction devices have, as a rule, a higher current gain (up to 20 db) than point contact transistors. Finally, a junction transistor can have a much lower noise level than a point contact transistor.

While, in point contact devices, the minority carriers increase the reverse current, in junction devices they also enhance the forward current. The proportion of current entering the base is high in point contact and low in junction devices. On this account junction, much

more than point contact, transistors resemble the vacuum triode with its very small grid current.

The analyses of all these details and of the properties of different types of transistor, together with the problem of their applications in definite circuits, form the subjects of special studies.

It should be pointed out that semiconductor diodes and transistors are at present in the process of rapid development and improvement. Other semiconductors are being studied, besides silicon and germanium. Other phenomena, apart from those associated with *p–n–p* and *n–p–n* structures are being investigated, especially modulation by high fields. Tetrode structures have followed diodes and triodes. New applications, which cannot be fitted by vacuum valves, are being discovered for semiconductor devices.[31, 32, 33]

Transformation of light and nuclear radiation into electrical energy by means of *p–n* junctions

The equilibrium between two sides of a *p–n* junction is maintained by the potential difference existing between them, which prevents the escape of electrons from the *n*-type, and holes from the *p*-type, material. The same potential difference draws into the *n*-type region any electrons appearing near the junction and, likewise, any holes are drawn into the *p*-type region. These free carriers may be generated near the junction by the absorption of photons or nuclear radiation. The basic process of the internal photo-effect consists in the formation of a hole–electron pair through the excitation of an electron from the filled band, across the forbidden gap, into the conduction band. In this process the frequency ν of the incident radiation must exceed the red limit of absorption, i.e. we must have $h\nu > \Delta E_0$, where ΔE_0 is the width of the forbidden gap. Depending upon the conditions of recombination of holes and electrons, one or the other carrier may contribute predominantly to the conduction process with a continuous illumination. As a rule, the sign of photo-conductive effect is the same as for conduction in the dark.

However, this condition is reached only as a result of preferential trapping of one type of carrier. The primary result of the absorption of a photon is the simultaneous creation of a hole–electron pair. An electron appears as a minority carrier on the *p*-type side of a *p–n* junction, while a hole has the same role on the *n*-type side. Both these carriers, diffusing towards the common boundary, are accelerated there and acquire an energy Ve, where V is the potential difference existing at the *p–n* junction during illumination. V is less than the potential difference V_0 appearing in the dark. V_0 corresponds to the difference between the distances of chemical potential from the conduction band in *p*- and *n*-type regions.

V is determined by the voltage–current characteristic, which gives the relation between the external potential on a p–n junction and the density j of hole and electron currents. V is equal to V_0 only under the open-circuit conditions.

It is essential, for the absorption of a photon and the creation of a hole–electron pair, that the energy $h\nu$ of the photon should exceed ΔE_0.

Assuming that each photon creates a hole–electron pair and that all pairs reach the p–n junction, the efficiency of the photoelement, i.e. the ratio of electrical energy liberated at the element to the energy of the absorbed light is $eV/h\nu$. This ratio might be of the order of some tens per cent. It should be borne in mind, however, that the spectral composition of solar radiation as well as of radiation from artificial sources comprises a wide range of frequencies, a large part of which may not satisfy the condition $h\nu > \Delta E_0$. Furthermore, for high frequency radiation such that $h\nu \gg \Delta E_0$, the quantum efficiency, i.e. the ratio of the number of electrons liberated in the semiconductor to the number of absorbed photons, is considerably less than unity. Both these factors lower the efficiency of the photocell.

The efficiency, with solar radiation, of the best present-day silicon photocells is 11%, and that of germanium photocells is 3–5%. The efficiency of the first photocells using cuprous oxide did not exceed 0·1%, while the best pre-war photocells, using thallium sulphide or silver sulphide, had an efficiency of up to 1%, for solar radiation. The efficiency of the photo-chemical reaction, corresponding to photosynthesis in the leaves of living plants, also does not exceed 1%. A reversible photo-chemical reaction using solar radiation, if it were feasible, could have an efficiency of up to 95%.

We shall consider a third method of utilizing solar energy, i.e. by means of thermoelements. Using light focused by means of mirrors to achieve a working temperature of the thermoelements of 300–400° C, and removing heat from the cold side at a temperature 50–100° C, it is possible to attain an efficiency of 5–10%.

Notwithstanding the high cost of mirrors, especially with provision for their movement following the sun, the thermoelectric method is at present about a hundred times cheaper than the photoelectric method. 1 cm^2 of a silicon photoelement costs about 4 dollars. An area in excess of 10 m^2 would have to be covered in order to obtain 1 kW of electric power, thus requiring an expenditure of some 400,000 dollars. A thermoelectric plant would require a capital outlay of approximately 1000 dollars per kW of electrical energy.

Small-scale solar generators, based both on the thermoelectric and on the photoelectric principle, have been constructed and are operating.

The application of *p–n* junctions to the conversion of the energy of nuclear radiation and X-rays into electrical energy is largely similar to the conversion of light energy and can be reduced to the generation of hole electron pairs on both sides of a *p–n* junction. However, the energy of particles (electrons, photons or nucleons) in this case exceeds by large factors the limiting energy ΔE_0 necessary for the liberation of a hole–electron pair. The excess energy is imparted to the liberated charges which in turn give rise to the formation of hole–electron pairs with only a slightly reduced kinetic energy. This process of pair creation and the successive reduction of energy continues until the energy of the liberated carriers is insufficient to cause further dissociation of electrons.

As a result of this process, a large number of electrons (up to 200,000 in the case of radioactive strontium) with an energy not exceeding ΔE_0, are created by every particle having an energy of the order of some hundreds of thousands eV.

The effectiveness of such devices depends upon the ratio of the total energy of all charges created in a distance not exceeding a diffusion length from the *p–n* junction to the energy of the primary particle. Analysis shows that the efficiency of energy transformation may, under favourable conditions, reach 40%.

A drawback of this method consists in the appearance of secondary processes due to the passage of high-energy particles through the medium.

Transmutations and displacements of atoms from their equilibrium lattice sites are caused by particles of sufficiently high energy. The density of defects increases with the passage of time changing the properties of the semiconductor, which then becomes ineffective. Experiments have shown that electrons with energies below 200,000 eV are relatively harmless in the case of silicon.

A semiconductor containing a *p–n* junction and coated with a radioactive substance represents a source of energy. The life of this source increases with the half-life of the radioactive material but, at the same time, the radiation intensity and power decrease. For some applications requiring only a low power supply, such as a transistorized radio receiver, a radiation battery of this type could provide a source lasting some years.

A semiconductor with a *p–n* junction may be used as a dosimeter for radiation; in this case the measurement, rather than the supply, of energy is involved.

Emission of charge carriers from a semiconductor

The electrons and ions present in a semiconductor may leave the boundaries of the material and enter the surrounding space if given

sufficient kinetic energy. This may be accomplished through the increase of temperature, the absorption of light or the impact of high-energy electrons.

The energy required by an electron in order to traverse the external boundary of a solid is a measure of its binding energy to the solid. This is only valid, however, in the case where none of the remaining charges change their state during the process of the removal of electrons. This condition is sufficiently well satisfied in metals which have a high electrical conductivity. The electric field due to an emitted electron is effectively compensated by a transfer of conduction electrons in the metal without appreciably altering the distribution of the remaining electrons. Different conditions obtain in a semiconductor with a low concentration of free carriers. In this case the mechanism of emission of electrons from the structure of the material must be considered very carefully.

On traversing the external boundary of an initially neutral semiconductor, the electron leaves behind a positive charge which polarizes the surrounding medium. Also, the field of the receding electron penetrates into the bulk of the semiconductor.

In a semiconductor of low conductivity, the carrier distribution cannot be altered appreciably in the short time required by the emitted electron to traverse a distance large compared with interatomic distances, say in excess of 10^{-6} cm. The interaction between the receding electron and the semiconductor is effectively due to the attraction between the former and the positive charge it left behind at the surface. The work expended in overcoming this attraction is

$$\int_a^\infty \frac{e^2}{r^2}\,dr = \frac{e^2}{a},$$

where a is the minimum distance at which the electron may be regarded as being outside the solid, leaving behind an excess of positive charge.

The magnitude of a may be regarded as being of the order of interatomic distances in the solid—about 3×10^{-8} cm. The work function expressed in electron-volts is given by:

$$\frac{e^2}{a} = e\phi$$

and thus:

$$\phi = \frac{e}{a} = \frac{5 \times 10^{-10} \times 300}{3 \times 10^{-8}} = 5 \text{ volt.}$$

This calculation assumes that the electron has been removed from the very surface layer of a semiconductor which is neutral throughout its volume. If the removed electron has come from a certain depth, the result is different.

Assuming that the emitted electron has a velocity of the order 10^7 cm/sec and that the maximum distance at which it still experiences an appreciable attraction is of the order 10^{-6} cm, the time required to travel that distance is found to be 10^{-13} sec. In order to change the field induced by this electron in a layer of the semiconductor, of thickness 10^{-6} cm, by a factor e, in this time, it is necessary to have a conductivity in excess of 10^{-5} ohm^{-1} cm^{-1}. If the conductivity is considerably in excess of this value, the electric field due to the emitted electron is at all times compensated by the charges in the conductor. The distribution of positive charges formed near the surface of the conductor is, in this case, such that their lines of force are symmetrical to the field of the electron, but with opposite sign (Fig. 95). The attractive force between the electron and its image in the conductor is equal to $e^2/4r^2$ and the work of removing the electron from a distance a to infinity is:

$$e\phi_1 = \int_a^\infty \frac{e^2}{4r^2}\,dr = \frac{1}{4}\frac{e^2}{a}.$$

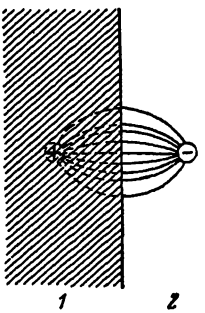


Fig. 95. Lines of electric field due to an electron emitted from a metal (1) into a vacuum (2).

Furthermore, calculation shows that the same amount of energy is required to remove the electron from the surface to the distance a beyond which the attraction may be regarded as being due to the image force. Thus, the total work required to remove an electron from the surface of a good conductor is

$$e\phi_2 = \frac{1}{2}\frac{e^2}{a}.$$

In most cases, however, the electron does not originate at the surface but at some depth of the order of the free path, i.e. about 5×10^{-7} cm. In this case, the work done against the force of attraction between the electron and the positive charge left behind in the semiconductor is

found to be small. This is due, firstly, to the fact that the distance between the emitted and the remaining charges is about 20 times larger and, secondly, to the fact that the field of the remaining charge is confined to a medium with a high dielectric constant, of the order of 10. Thus, the work of removing the electron from the residual positive charge is reduced by two orders of magnitude to a few hundredths of an eV.

However, the receding electron polarizes the semiconductor by its field and its removal from this polarization field may be calculated on the assumption that the semiconductor represents a continuous medium with a dielectric constant ϵ.

An electron at the surface is surrounded by an electric field, one half of which is situated in the semiconductor, while the other half is in the surrounding medium for which we assume the dielectric constant to be unity. On leaving the medium, however, the electron is surrounded on all sides by a field situated in the medium with unit dielectric constant. Thus, the change of energy of the field surrounding the electron amounts to:

$$e\phi = \frac{1}{2} \cdot \frac{e^2}{a} \cdot \frac{\epsilon - 1}{\epsilon + 1}.$$

In this expression ϵ refers to the magnitude of dielectric constant which appears as a result of the polarization of the medium during the time of motion of the electron, i.e. 10^{-13} sec. It is necessary to distinguish polarization due to two causes: (1) the displacement of electrons in the atoms of the material, taking place in a time 10^{-15} sec or less (this part of the dielectric constant is equal to the square of the refractive index for long waves); (2) the displacement of ions due to the electric field taking place in a time of the order of the period of their vibrations, i.e. 10^{-12}–10^{-13} sec.

In the case under consideration, we need only take into account the work done against that part of the polarization which is due to electrons, since the polarization due to ions cannot change in the short time of interaction with the electron flying away from the surface. This part of the polarization will decay later, releasing its stored energy and transforming it into heat.

Under these circumstances, where the electronic part of the dielectric constant does not exceed 4, the relevant part of the work function, for a equal to 3×10^{-8} cm, does not exceed:

$$\frac{1}{2} \cdot \frac{4-1}{4+1} \cdot \frac{e^2}{a} = 0{\cdot}3 \frac{e^2}{a}.$$

In metals, at the absolute zero, the electrons have energies ranging from zero to the limiting Fermi energy, which coincides with the

chemical potential μ. In order to be able to leave the metal the electron should have an additional energy not less than the work function, $e\phi$, so that the total kinetic energy K of the emitted electron, if it moves normal to the surface, should exceed $(\mu + e\phi)$. This quantity characterizes the binding energy of the electron in the metal. After the emission of the electron, the remaining electrons in the metal retain their equilibrium Fermi distribution. Fig. 96 illustrates the work function of an electron

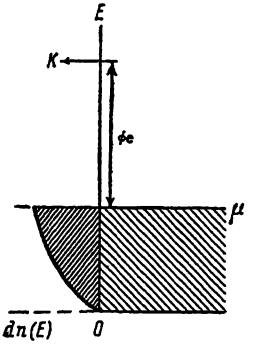


Fig. 96. The work function of electrons in a metal.

in a metal. The ordinates denote electronic energies and the abscissae their concentrations at different energies. The arrows represent the excess energies which must be acquired by the electron in order to enable it to leave the solid. The density of levels increases as the square root of the energy on approaching the Fermi level and the energy required for emission from the solid decreases. Thus, most of the electrons emitted from the metal come from levels close to the Fermi level.

The conditions of electron emission from a semiconductor differ appreciably from those obtaining in metals. There are no electrons at the chemical potential level μ at which the work function is ϕe. The electrons which may leave the interior are either in the valence band $(\phi_3 e, \phi_4 e)$ (Fig. 97), or in the conduction band $(\phi_1 e)$, or in the impurity levels $(\phi_2 e)$. The majority of valence electrons are in the valence band and, thus, the most probable transitions are from levels of lower energies. The density of electrons in a given energy interval decreases as we approach the upper edge of the valence band and so does the probability of their emission due to external causes such as light or impact of fast electrons. Consider the case of emission under the influence of absorption or collisions of a photon, which imparts to the electron an energy exceeding the minimum required for the overcoming of the external energy barrier. In the case of a metal the emitted electrons will be

derived predominantly from levels near the Fermi level. In a semiconductor, on the other hand, they will be drawn predominantly from the valence band, and then not only from levels near the top of this band.

In a semiconductor in contact with a metal, the chemical potential μ coincides with the Fermi level, but the position of the edge E_2 of the valence band, which is the chief source of electrons, depends to a large extent upon the presence of impurity levels in the forbidden band. The energy level of the top of the valence band in a pure material (intrinsic semiconductor) falls approximately $(E_1-E_2)/2$ below μ, where (E_1-E_2) is the width of the forbidden band.

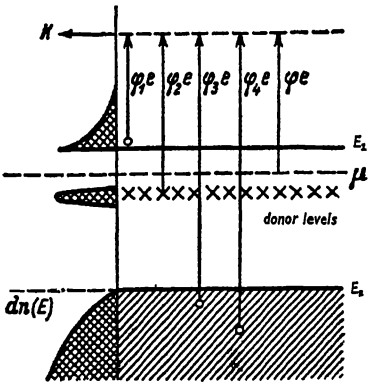


Fig. 97. The work function of electrons in an n-type semiconductor.

In a semiconductor, with donor levels which act as a source of electrons, the distance between μ and E_2 increases, while in a p-type semiconductor with acceptor levels this distance decreases. Furthermore, the position of the chemical potential in the energy band system of the semiconductor changes somewhat with temperature.

The energy which must be imparted to an electron in order to remove it from the semiconductor depends, therefore, upon whether the electron comes from the valence band, the conduction band or the impurity levels. The energy difference which is established between a semiconductor and a metal in contact with it also depends upon the position of the chemical potential in the energy level system of the semiconductor.

In the neighbourhood of the absolute zero the minimum energy, which must be imparted to an electron in order to remove it from the metal, is equal to the energy difference between the Fermi level and the top of the potential well of the metal. On raising the temperature, the Fermi distribution function becomes smeared out over a distance of the order kT, i.e. over a few hundredths or tenths of an eV.

In a semiconductor, only a relatively small part of the electrons, those present in the conduction band, require the minimum excess energy

in order to leave the solid. A somewhat higher energy is required by electrons occupying the impurity levels, while the maximum energy is necessary for the removal of valence band electrons.

All these considerations are illustrated in Fig. 98 which refers to the contact between a metal and an n-type semiconductor.

After an electron has left the semiconductor, the thermal motion of the remaining electrons restores their statistical distribution. Thus, for example, if a conduction electron having a kinetic energy greater than

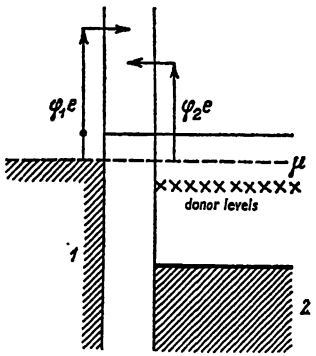


Fig. 98. The work function and chemical potential level of a metal (1) and an n-type semiconductor (2).

$\phi_1 e$ leaves the semiconductor, its place is taken by an electron coming from the impurity levels. Both before and after the emission of an electron, the electron distribution among energy levels is determined by the position of these levels with respect to the chemical potential μ. Therefore, the excess free energy required for emission of an electron from the semiconductor is equal to ϕe.

We should distinguish between an adiabatic process of emission of a free electron which requires an energy $\phi_1 e$ but leaves the solid in a non-equilibrium state, and an isothermal process requiring an energy ϕe. Under thermal equilibrium conditions, the thermal emission of electrons is determined by the difference between the energy K outside the semiconductor and at the level μ.

The work function ϕe depends, in common with all other statistical properties of a semiconductor (concentration of free electrons and holes, thermoelectric power, etc.) upon the position of the chemical potential. It may have different values in the same material, depending upon the temperature, impurity concentration and external pressure.

One parameter which is characteristic of the basic material itself and does not depend upon the position of the chemical potential is the minimum kinetic energy $\phi_1 e = (K - E_1)$, or $\phi_2 e = (K - E_2)$, which is lost in the process of adiabatic emission of an electron from the

semiconductor. The magnitudes of ϕ_1 and ϕ_2 are determined by the crystal lattice and not by the impurities, in the same way as the width of the forbidden band, the density or the dielectric constant.

Apart from these general considerations we may consider those physical properties which are determined by the emission of electrons from solids. These are the magnitude of the contact potential, the external photoeffect, secondary emission and emission of electrons under ionic bombardment.

The contact potential

The contact potential of both metals and semiconductors is determined by the difference ϕe between the energy outside the conductor and the chemical potential level (Fig. 99).

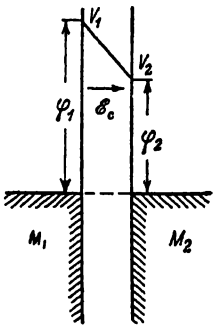


Fig. 99. Contact potential difference at the interface between two metals M_1 and M_2.

Under thermal equilibrium conditions, the chemical potentials of all solids in contact are equal. On the other hand, the outer boundary of each of these bodies acquires a potential corresponding to the energy of electrons which are just prevented from leaving the solid. As this potential varies from one material to another, depending upon its binding energy for electrons, the surface potentials of different materials in contact differ between themselves. A contact potential field $\mathscr{E}_c$ appears between them as well as a contact potential difference V_c equal to $(V_1 - V_2)$ (Fig. 99).

Different methods may be employed in the measurement of contact potential difference and they all give good agreement in their results. The simplest method is based on compensation and it consists in the following procedure. The two surfaces under investigation are connected in series with a variable potential difference which is so adjusted as to compensate exactly for the measured contact potential difference. The criterion of complete compensation is the absence of an electric field in the space between the two surfaces. If the distances between

these surfaces is varied periodically, thus varying the capacitance, there will be no induced charges only in the absence of an electric field. If the air in the space between the surfaces is ionized, current will cease to flow under conditions of compensation.

This method has also been applied to semiconductors. In particular A. V. Ioffe measured the contact potentials of a large number of metals and semiconductors in air using the second variant of the method.

TABLE 7

Material	No. of samples	Contact potential with respect to gold	Material	No. of samples	Contact potential with respect to gold
MoS_2	5	$+0{\cdot}47\pm0{\cdot}09$	Cu_2Te	2	$-0{\cdot}33\pm0{\cdot}01$
UO_2	2	$+0{\cdot}27\pm0{\cdot}03$	Tl_2S	4	$-0{\cdot}47\pm0{\cdot}02$
NiO	2	$+0{\cdot}27\pm0{\cdot}05$	Aquadag	4	$+0{\cdot}07\pm0{\cdot}03$
CuO	10	$+0{\cdot}23\pm0{\cdot}18$	Au	16	$0{\cdot}00\pm0{\cdot}07$
V_2O_5	3	$+0{\cdot}08\pm0{\cdot}01$	W	1	$-0{\cdot}03$
WO_3	5	$+0{\cdot}06\pm0{\cdot}07$	Graphitized rubber	3	$-0{\cdot}04\pm0{\cdot}04$
Cu_2O	11	$+0{\cdot}05\pm0{\cdot}04$	Cu	1	$-0{\cdot}04$
MnS	3	$+0{\cdot}05\pm0{\cdot}04$	Pt	1	$-0{\cdot}09$
CoS	1	$+0{\cdot}04$	Mo	3	$-0{\cdot}31\pm0{\cdot}04$
Se	4	$+0{\cdot}03\pm0{\cdot}02$	Ni	1	$-0{\cdot}31$
NiSe	1	$-0{\cdot}07$	Bi	1	$-0{\cdot}34$
Cu_2S	3	$-0{\cdot}08\pm0{\cdot}04$	Si	3	$-0{\cdot}46\pm0{\cdot}09$
MnTe	2	$-0{\cdot}12\pm0{\cdot}02$	Zn	2	$-0{\cdot}55\pm0{\cdot}09$
SnSe	4	$-0{\cdot}14\pm0{\cdot}06$	Al	14	$-0{\cdot}60\pm0{\cdot}10$
NiS	1	$-0{\cdot}15$	BiSnCd	3	$-0{\cdot}62\pm0{\cdot}02$
MnSe	2	$-0{\cdot}16\pm0{\cdot}06$	Cd	3	$-0{\cdot}80\pm0{\cdot}04$
ZnO	6	$-0{\cdot}17\pm0{\cdot}05$	Tl	1	$-0{\cdot}82$
CdS	5	$-0{\cdot}18\pm0{\cdot}07$	Be	1	$-1{\cdot}10$
CoTe	1	$-0{\cdot}21$	Mg	11	$-1{\cdot}2\pm0{\cdot}2$
TiO_2	10	$-0{\cdot}23\pm0{\cdot}08$			
PbS	2	$-0{\cdot}27\pm0{\cdot}04$			

Experiment showed that the contact potentials depend strongly upon the condition of the surface: the presence of adsorbed gases and, especially, adsorbed ions, oxide layers, and thin (one- to two-) atomic layers of any kind, including liquid layers. A sharp distinction should be made between the value of contact potential of a material with a carefully cleaned surface in a vacuum and the value obtained in the atmosphere of the laboratory. The contact potential in the latter case is not a constant of the material, but its value is of interest for the understanding of phenomena occurring at a surface exposed to air.

Table 7 gives the values of contact potential, for a number of semiconductors and metals in air, obtained by A. V. Ioffe.[34]

In accordance with expectations based on the foregoing considerations, the contact potential of *n*-type semiconductors are, as a rule, lower than those of *p*-type semiconductors. Experiment shows, further,

TABLE 8

Metal	ϕ (V)	λ_0 (Å)	Metal	ϕ (V)	λ_0 (Å)
Li	2·34–2·38	5280	Ge	4·55	2720
Na	2·33	5300	Sn (tetrahedral)	4.51	2740
K	2·26	5460	Pb	4·15	2980
Rb	2·13	5800	V	3·78	3270
Cs	1·93	6400	Ta	4·12–4·16	3000
Cu	4·29	—	As	5·23	2360
Ag	4·73	2610	Sb	4·02	3070
Au	4·76	2600	Bi	4·62	2680
Be	3·92	3150	Co	3·72	3320
Mg	3·69	3350	Mo	4·19–4·29	2990
Ca	2·96	4180	W	4·55–4·57	2720
Sr	2·25	5500	U	3·28	—
Ba	2·55	4840	Mn	3·76	3200
Zn	4·25	2910	Re	4·98–5·10	2480
Cd	4·08	3030	Fe	4·75–4·77	2600
Hg	4·53	2730	Co (hexagonal)	4·25	—
Al	4·25	2910	Ni	4·98–5·03	2490
Gd	3·00	4120	Rh	4·75	2600
Tl	3·68	3350	Pd	4·97–4·99	2490
Ti	3·92	3150	Pt	5·44–6·37	1900
Zr	4·13	—	Se	4·89	2530
Ce	3·07	—			
Th	3·29	—			

TABLE 8A

Compound or composite cathode	ϕ (V)	λ_0 (Å)	Compound or composite cathode	ϕ (V)	λ_0 (Å)
BaO	1·00	12350	Ba on W	1·66	—
CuO	5·34	2310	Ba on oxidized W	1·1	—
Cu_2O	5·15	2390	Cs on oxidized Ag	0·71	17400
Cs on W	1·36	9090	Th on W	2·62	—
K on Pt	1·62	7700			
BaO/SrO on Pt	1·03	12380			

also in accordance with expectations, that a surface layer of a foreign material deposited on a substrate gives the contact potential characteristic of the material of the layer and not of the substrate, even if the layer is only 3–4 atoms in thickness.

A monatomic or mono-molecular layer, on the other hand, does not give rise to its characteristic contact potential value although it does affect the contact potential of the substrate.

If the adsorbed molecules have a dipole moment, the corresponding potential difference is added to, or subtracted from, the contact potential of the substrate. The subsequent molecular layer is oriented in the opposite sense and cancels the influence of the first layer. The third layer again restores the potential to the original value.

A. V. Ioffe and Lukirskij have found periodic variations of the contact potential of gold due to successive depositions on its surface of one, two, three and more mono-molecular layers of fatty acids which caused corresponding alternate variations of the contact potential. The French physicist Trillat confirmed such mutual disposition of layers by means of X-ray analysis.

The deposition of a monatomic layer of potassium or caesium on the surface of a metal lowers the work function appreciably below the values of the contact potential for the alkali metals themselves. A particularly pronounced effect is obtained with mono-molecular layers of the oxides or sulphides of alkali, or alkali earth, metals, which are typical semi-conductors. The work function falls, in this case, from 4–5 V to 1 V and less.

In order to obtain large thermionic emission currents in a vacuum, it is important to secure the lowest possible values of the contact potential. These are obtained by deposition of monatomic layers of thorium or caesium on tungsten and by the use of thick layers of barium or strontium oxides with the addition of metallic barium or strontium. In many cases the monatomic layer covers only parts of the substrate surface and leads to the formation of active spots with lowered work function.

Tables 8 and 8a give values of the work function and the long wavelength limit λ_0 of the photo-effect for clean metals and semi-conductors with industrial applications.

External photo-effect

We have already considered the conditions under which electrons are emitted from metals and semiconductors, as a result of absorption of light. Photons of sufficiently high frequency may liberate electrons from the solid, provided they are absorbed within a distance of the order of the electronic free path from the surface of the material and that the electron velocity is approximately normal to the surface. The photo-electric properties of a solid thus depend not only upon the condition of the surface which determines the work function of the material, but, also, upon the conditions of light absorption within the solid.

Theoretical and experimental investigations have been made of a number of problems such as: the external photo-effect from clean metals, determination of the depth of emission of electrons, the velocity distribution of the emitted photo-electrons, the dependence of the photo-current upon the wavelength of the exciting light and the temperature of the material. In all cases the theoretical results have been well confirmed by experiments.

Semiconductors have been much less extensively studied from these points of view, although all modern materials for photo-cathodes belong to this class. The phenomenon of the external photo-effect in a semiconductor has a much more complicated characteristic than in a metal. It depends, in the first place, upon whether the absorption of light is caused by the electrons in the valence band or by impurities. The red limit may be shifted towards longer waves by the addition of impurities acting as sources of electrons. The impurity atoms may serve as sources of electrons if their concentration is high. Apart from this, in cases where the impurity levels are situated close to the conduction band, the conduction electrons may also take part in the process of photo-emission. Their density increases with temperature.

In the case of electron emission from the valence band, the increase of photo-current, with increasing frequency of the absorbed light, is determined by the variation of the density of states in the valence band as a function of the distance from the band edge. Let ν_0 denote the threshold frequency corresponding to the red limit. Assume also that only electrons within a distance kT from the top of the valence band may be emitted from the semiconductor at this frequency. At any higher frequency $\nu > \nu_0$, all the valence band electrons situated at distances less than $(\nu - \nu_0)/h$ from the band edge may contribute to the photo-current. The density of states increases with distance from the band edge as $(\nu - \nu_0)^{\frac{1}{2}}$. Consequently, the number of photo-electrons and, hence, the photo-current, should increase with increasing frequency as $\beta(\nu - \nu_0)^{\frac{1}{2}}$, where β is the probability of absorption of a photon by an electron.

Experimental work by Arsen'eva has confirmed such a frequency-dependence of the photo-current. A much more rapid increase of current has been found in metals. This corresponds to the fact that, in metals, the density of states reaches its maximum at the Fermi level and decreases slowly with decreasing energy.[35]

The actual value of ν_0 in semiconductors depends upon the position of the chemical potential μ which, in turn, depends upon the density and type of impurities.

The external photo-currents are not large in common semiconductors. It is possible that this may to some extent be due to the absorption of light. Strong absorption sets in as soon as the frequency of light becomes

sufficient for the excitation of electrons from the valence into the conduction band, i.e. when

$$h\nu > (E_1 - E_2)$$

where $(E_1 - E_2)$ is the width of the forbidden gap.

In order to traverse the surface and escape into the surrounding space, however, an electron requires a much larger amount of energy, viz. $e\phi_4 \gg (E_1 - E_2)$.

Thus, the external photo-effect requires light of a frequency ν_{ext} which is much higher than the frequency necessary for absorption of light and the internal photo-effect. At these frequencies light is very strongly reflected from the surface and only a very small fraction penetrates into the semiconductor and gives rise to the external photo-effect. Experiments show that the internal photo-effect, i.e. photoconduction, falls off rapidly in the region of strong absorption of light. This leads us to expect that the external photo-effect, resulting from those electrons liberated by light which have a normal component of velocity corresponding to energy exceeding $e\phi$, is also small.

Of technical importance are caesium oxide and antimony–caesium photo-cathodes, first suggested by Lukirskij and studied in detail by Lukyanov. The former consist of a substrate of silver covered with silver oxide, having a surface layer of caesium oxide and additional caesium atoms. Atoms and small aggregates of metallic silver and caesium are distributed in the photo-sensitized layer of the semiconductor (caesium oxide). Electrons from these impurities provide the main part of the photo-current. A small fraction of the photo-current is found, however, in the far infra-red region of the spectrum, at wavelengths of the order of 1·5 μ, and this cannot be ascribed to electrons from these impurities. It seems that this effect is due to electrons from the conduction band.

The antimony–caesium photo-cathodes show a highly selective spectral response: the photo-current reaches a maximum in the green region of the spectrum and falls off rapidly with frequency on either side of it. The quantum efficiency of an antimony–caesium photo-cathode in the frequency range corresponding to the maximum of the photo-current is exceptionally high. The number of electrons reaches 25–30% of the number of photons falling on the surface.

By comparison a caesium oxide photo-cathode has a quantum efficiency of 1% in its sensitive range (which does not exhibit any pronounced peak). Other photo-cathodes have a still lower efficiency.

The selective photo-effect of antimony–caesium photo-cathodes can be explained in terms of selective absorption of light by electrons on

impurity levels. I found a similar absorption, as early as 1904, in X-ray-activated rock salt and sylvite. It has subsequently been observed by Pohl and called "absorption by *F*–centres".

The selective photo-effect found in colloidal films of alkali, or alkali earth, metals has, however, a purely optical origin. The density of radiant energy at the depth which contributes the major part of electrons emitted outwards has a maximum at a given wavelength. This gives rise to selective sensitivity in the relevant spectral range. The number of electrons which, in the course of photon absorption, acquire a component of velocity normal to the surface depends upon the orientation of the plane of polarization of the light; the photo-current varies accordingly.

We, therefore, come to the conclusion that in the selective photo-effect in alkali metals the quantum efficiency remains a monotonic function of wavelength. On the other hand, the intensity of light, at that depth which gives the main contribution to the number of electrons has a maximum in the selective region.

In fact, Lukirskij and Khurgin have computed the square of the amplitude of the light wave, at the corresponding distance from the surface, for a number of metals (potassium, caesium, platinum). Their calculations show that the ratio of photo-current to the intensity of light wave at this depth varies with wavelength, but does not show a selective maximum at the point where the photo-current itself has a maximum. In all these cases the photo-current is determined by the number of photons absorbed in a thin layer near the surface.

The photo-current is increased by the application of a strong external electric field to the photo-cathode. This field accelerates afresh the electrons which have been slowed down by the action of the contact field of the surface. The external field superimposed on the contact field lowers the height of the energy barrier. At the same time, by reducing the barrier thickness, the field renders the barrier partly transparent to electrons with energies less than the barrier height. As may be expected, the influence of an external field is especially pronounced near the red limit of the photo-effect, where there are numerous electrons with energies which are otherwise not quite sufficient to overcome the potential barrier.

Thermionic emission

At sufficiently high temperatures, a substantial number of the conduction band electrons acquire kinetic energies which enable them to overcome the energy barrier and leave the boundary of the semiconductor. According to equation (84), the number n of electrons having such an energy (disregarding the direction of their motion) is

given by

$$n = \frac{2(2\pi m^* kT)^{\frac{3}{2}}}{h^3} \exp\left(-\frac{E-\mu}{kT}\right)$$

where $(E-\mu)$ is the difference between the energy required for emission and the chemical potential level μ. The kinetic energy of electrons having a total energy E is $(E-E_1)$, where E_1 is the energy at the conduction band edge.

However, not all the electrons having energies in excess of E can leave the semiconductor; only those can do so which, after their last

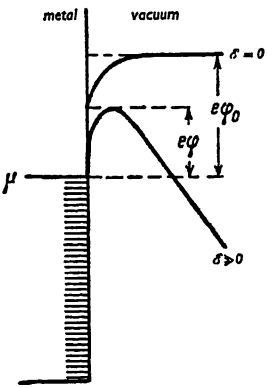


Fig. 100. The effect of an applied electric field upon the emission of electrons from the surface of a solid into vacuum.

collision, have a normal component of kinetic energy exceeding $(E-E_1)$. Even these electrons cannot all traverse the energy barrier: some electrons having a sufficient energy are reflected. In the absence of an accelerating electric field outside the barrier, the latter may be regarded as being infinitely wide, having a height E. The barrier, therefore, appears as non-transparent to all electrons with normal components of kinetic energy less than $(E-E_1)$.

Like the external photo-effect, thermionic emission increases in the presence, outside the radiating surface, of an electric field accelerating the electrons. This phenomenon is especially pronounced at temperatures for which the emission of electrons is only just apparent.

In the presence of a sufficiently high electric field $\mathscr{E}$, of the order 10^5–10^6 V/cm, the narrowing and lowering of the potential barrier ($e\phi$ instead of $e\phi_0$, see Fig. 100) assumes such proportions that an appreciable electron emission from the solid occurs, even at room temperature, by virtue of the tunnel effect. This phenomenon is referred to as cold emission. It may easily be observed, in a good vacuum, on charged sharp points which have a concentration of electric lines of force on their surfaces.

On the assumption of zero initial velocities for the emitted electrons, we find that a negative space charge is formed close to the emitting surface in the accelerating electric field. The field of this space charge reduces the external field in the neighbourhood of the surface and increases it outside the space charge layer. With parallel plane electrodes, at a distance d apart and with a potential difference V between them (including the contact potential difference), the maximum current j which can flow under these circumstances is

$$j = 1{\cdot}27 \times 10^{-9} V^{\frac{3}{2}}/d^2 \text{ A/cm}^2 \quad . \quad (153)$$

if V is expressed in volts and d in cm. According to this expression, the current becomes zero if V is zero, since we have neglected the initial velocities of electrons.

Bursian has analysed the problem of thermionic emission, taking into account a statistical distribution of velocities among the emitted electrons. The results cannot be given in such a simple form as for the case of zero initial velocities. It is possible, however, to give for every specific case the distribution of potential and of space charge and the density of the resulting current.

The thermionic electron current increases rapidly with the presence of positive ions in the region occupied by the negative space charge. With a sufficient concentration of these ions it is possible to compensate for the electronic space charge almost completely and to obtain currents which, even at low fields, transport all electrons emitted from the hot cathode.

The temperature at which thermionic emission becomes appreciable is the higher, the larger the magnitude of the work function. Thus, pure tungsten gives an appreciable current only at 1800° K, while thoriated tungsten, with a lower work function, gives the same current at 1000° K and tungsten covered with a monatomic layer of caesium at 600° K. The work function of pure tungsten is 4·5 eV and that of thoriated tungsten 2 eV. The work function of a layer of caesium on tungsten is 1·5 eV, while that of a caesium layer on oxidized tungsten is only 0·6 eV.

The oxides of the alkali earth metals were applied to the generation of thermionic currents at relatively low temperatures as early as the first decade of this century: at first calcium oxide and, next, a mixture of barium and strontium oxides with an excess of metallic barium and strontium, were used. The work function of such a semiconducting emitter does not exceed 0·6 eV. By determining the sign of thermoelectric power, Morgulis has shown that the excess barium atoms impart to the oxide an n-type conductivity. The thermionic emission increases with increasing excess of barium.

Secondary electron emission

A beam of fast electrons, which is absorbed by a solid, gives rise to electron emission from this solid. The coefficient of secondary emission σ is defined as the ratio of the electron current leaving the solid to the current striking the surface. This coefficient depends upon the nature of the emitter and upon the energy of the primary electrons. Secondary emission may be obtained both from metals and from semiconductors of different conductivities. The values of the coefficient of secondary emission σ of some materials are listed in Table 9.

TABLE 9
Coefficients of Secondary Emission

Compound	σ	Compound	σ	Compound	σ
WS_2	1·04	BaO	4·8–6·5	MgO	17·5
MoS_2	1·10	NaI	5·5	Ag_2O, Ni	9·0
Ag_2O	1·20	LiF	5·6	BeO, Ni	2·0
Cu_2O	1·25	KI	5·6	KI, Cs	12·6
MoO_2	1·25	RbCl	5·8	CsCl, Cs	13·7
CaF_2	3·15	NaCl	6·8	Ag_2S, Cs	17·5
MgF_2	4·10	Cs_2O, Ag	7·2	NaCl, Cs	22·7

The coefficient of secondary emission of metals usually does not exceed unity, even for the primary electron energies, of the order 500–1000 eV, which give maximum emission. In this case the energies of the secondary electrons lie in the range 0–10 eV.

Semiconductors give much higher values of the coefficients, as high as 10 or more. The majority of the secondary electrons have energies in the range 0–3 eV, although some reach energies as high as 10 eV.

It is found that the introduction, on the surface of a metal with a coefficient σ_M, of a thin layer of a semiconductor with a higher value σ_P, gives rise to gradual increase of σ from σ_M to σ_P with increasing layer thickness, the upper limit being reached at about $0{\cdot}1\mu$ (10^{-5} cm) layer thickness. Further increase of thickness produces no increase of σ.

The coefficient of secondary emission increases with increasing angle of incidence of the primary electrons. For oblique incidence at an angle of 80°, σ is larger by a factor of almost 2 compared with the value for normal incidence.

This dependence of σ upon the thickness of the semiconductor layer and the angle of incidence suggest that the process of emission of secondary electrons takes place in the bulk of the semiconductor. The effective thickness of the surface layer amounts to some tens of free path lengths. It appears that an electron liberated at this depth acquires appreciable kinetic energy, which is subsequently dissipated in a series

of collisions with thermal lattice vibrations and other imperfections. This explanation is quite natural, since the most probable amount of energy lost in a single collision corresponds to the energy of a phonon of some hundredths of an electron-volt. An electron may suffer a large number of collisions and yet retain sufficient energy to overcome the potential barrier. A still higher number of collisions is necessary to bring a fast electron into statistical equilibrium corresponding to a given temperature.

The number of secondary electrons, created in a given distance Δl of the path of the primary electron, increases with decreasing velocity of the latter. Most secondary electrons are, therefore, liberated at such a depth that the primary electron has already been slowed down.

The magnitude of the work function does not have any large effect on the coefficient of secondary emission since the majority of electrons leaving a semiconductor, and more especially a metal, have high energies. Frequently, however, the best emitters are those materials which give the highest thermionic emission, such as oxide cathodes.

The coefficients of secondary emission of metals are almost independent of a rise in temperature. In semiconductors, on the other hand, this dependence is very pronounced on account, firstly, of the increasing number of free electrons which interact with the fast electrons and acquire energy from them. Secondly, there is another more trivial reason for this increase; an increased conductivity reduces the potential which is formed on the surface of a semiconductor under the influence of fast electron bombardment. Let j denote the density of the primary electron current, ρ the resistivity of the semiconducting layer and d its thickness. In the absence of secondary emission the potential difference V between the outer surface and the metallic base is

$$V = j\rho d.$$

Consequently, the initial energy $V_0 e$ of an electron is reduced to $(V_0 - V)e$ by the time it has reached the surface of the semiconductor at normal incidence. At oblique incidence, at an angle α, the normal component of velocity $(2eV_0/m^*)^{\frac{1}{2}} \cos\alpha$ is reduced by $(2eV/m^*)^{\frac{1}{2}}$. At a suitable angle α the incident beam may be partly reflected from the charged surface.

These considerations are modified if we take into account the secondary emission: even in the best insulating dielectrics, the surface potential adjusts itself so that σ is unity and the current j is zero.

In semiconductors the coefficient of secondary emission increases exponentially with temperature. According to Morgulis, the value of σ for an oxide cathode increases by a factor 5–6 from room temperature to 600° C. Khlebnikov has observed an increase of σ by $1\frac{1}{2}$ times raising the temperature of an antimony–caesium photo-cathode to 100° C.

Secondary emission has acquired considerable practical importance in view of the possibility of the high amplification of weak electron currents, which presents itself when σ appreciably exceeds unity (for a number of semiconductors the values of σ lie between 3 and 22). Kubetskij first realized this idea in his photo-multiplier.[36] In his instrument, shown in Fig. 101, electrons are liberated in small numbers

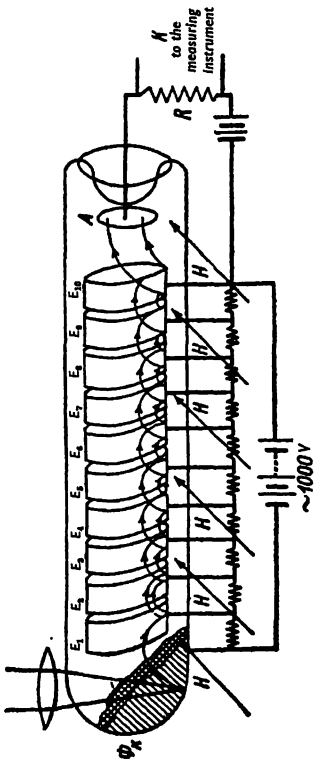


Fig. 101. Photo-multiplier of Kubetskij. ϕ_K—photo-cathode; E_1–E_{10}—accelerating electrodes with a high coefficient of secondary emission; A—anode; H—magnetic field.

at a sensitized photo-cathode under the action of light. These electrons, accelerated by an electric field and deflected by a magnetic field, fall on a layer of secondary electron emitter and their current is multiplied by a factor σ. By repeating this process a number of times, Kubetskij obtained a current j equal to $j_0 \sigma^n$, where j_0 is the initial photo-current and the exponent n is the number of secondary multipliers. As a result, currents of some 10^{-9} A were amplified to 0·5 mA. An amplification of this magnitude enables the measurement of exceedingly small currents,

even to the point of detecting individual electrons. The sensitivity of modern photo-multipliers reaches 40 A/lumen.

Another type of secondary electron emission, for which σ may reach values in excess of 100, can be observed in thin layers of semiconductors, the conductivity of which approximates to that of dielectrics. This type of emission, in contrast with the former type described above, shows a high inertia and frequently persists after the removal of the primary electron beam.

This particular type of emission is found in materials having $\sigma \gg 1$; their surfaces are charged positively as a result of bombardment with negative electrons. The positive surface charge gives rise to a very high field, of the order 10^6 V/cm, in the thin dielectric layer deposited on a metal substrate. The presence of this electric field, in a material of a high dielectric constant, causes cold emission from the substrate and increased conductivity. An electron acquires more energy in a free path length from the field than it can dissipate in collisions and, thus, accumulates energy as it moves across the dielectric layer. Eventually this energy may be sufficient for the escape of the electron. In fact, assuming a free path length of 5×10^{-7} cm and a field of 10^6 V/cm, the energy acquired between collisions is 0·5 eV, while the dissipation of energy in collisions takes place mainly in quantities of the order 0·03 eV.

In this process the field is due to the surface charge and emission goes on until the surface charge is reduced to sufficiently small values for the acceleration of electrons to cease. The removal of the surface charge, by any method, also removes this inertial emission.

Electron emission due to ions

The emission of slow electrons under the influence of ion bombardment was first explained by Kapitza in terms of the local heating which gives rise to thermionic emission. This explanation is supported by the fact that the secondary electrons, in this case, have a low velocity (as opposed to the cases of secondary electron emission and the photo-effect), corresponding to a local temperature of some 10,000° C.

The coefficients of secondary emission σ on ion impact are commonly lower than for electrons but may reach values of a few units. The magnitude of σ attains a maximum with an ion energy of the order of some hundreds of eV. The majority of secondary electrons are liberated by ions at a certain depth below the surface, as in the previously described case of emission under electron impact, but the depth is less in the present case.

Apart from Kapitza's theory, other hypotheses have been advanced, relating the mechanism of emission to the ionization of atoms by ion impact or to impact-excitation of an atom and transfer of energy to the electron.

Emission of electrons is also observed under the influence of bombardment of a material by fast neutral atoms. The coefficient of emission is somewhat lower than for positive ions of the same kinetic energy. For negative ions the coefficient is higher than for positive ones, while for electrons it is higher still.

These phenomena have been very little studied, especially in semiconductors. The basic work may be attributed to Starodubtsev and his co-workers.[37]

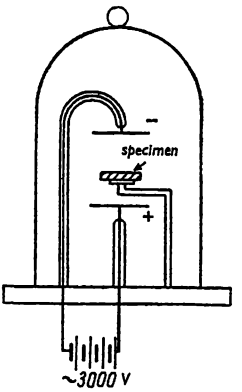


Fig. 102. Diagram of the apparatus for cathodic sputtering of metals.

Bombardment with charged particles may also give rise to the emission of ions (especially from polar crystals), neutral atoms and, sometimes, even small fragments of the material. All these phenomena are apparently due to the elevated temperature of the body or its parts and subsequent dissociation or evaporation.

The velocities of emitted particles correspond to a temperature of the order of 10,000° C and their directions do not depend upon the direction of the primary beam of fast particles. The ratio of the number of particles emitted from the surface, and especially their energy, to the number and energy of the primary particles, usually amounts to a few per cent in the case of metals, and to few tens per cent in the case of semiconductors. Much higher values of secondary emission coefficients for ions or atoms may be obtained in atmospheres of some inert gases at reduced pressures, but even then the values seldom exceed unity.

Cathodic sputtering has acquired a considerable practical significance. The atoms of the cathode in a gas discharge (Fig. 102) are sputtered into the surrounding medium, under the influence of bombardment by positive ions, and are deposited on the surrounding surfaces.

Cathodic sputtering is employed for deposition of thin metallic films on surfaces of various materials.

Contact electrification

Electrification by friction is the first known phenomenon of electrical interaction having been discovered over 2000 years ago. Amber, ebonite, glass, fur, and other insulating materials, under the influence of friction, acquire charges and attract small light objects, such as pieces of paper. Subsequent work has shown that the electrification of specimens is due, not to the energy dissipated in friction, but to the actual contact between two different materials, one of which acquires excess positive charge, the other excess negative charge. Rubbing only serves the purpose of bringing into contact the maximum possible number of surface elements.

The physical nature of the processes involved in electrification by friction is not clear to this day, notwithstanding the fact that this was one of the earliest means of obtaining and studying electricity and that a large volume of work has been devoted to it.

Gezekhus discovered and studied the effect of surface layer thickness, and its electrical and mechanical properties, upon the magnitude and sign of electrification. As a result of the analysis of a large number of data, Cohen formulated the rule that, of two substances in contact, the one with the higher dielectric constant acquires a positive charge. At first sight, this relationship is contrary to expectations. According to Gezekhus's interpretation, the charging of the surface is due to transition of electrons from the material in which their binding is weaker into the material in which it is stronger and the energy lower. One would, thus, expect electron transitions from the material with the lower dielectric constant to that with the higher dielectric constant, i.e. the latter would acquire negative charge. Yet Cohen found a predominantly positive charge on it.

Cohen's relation can be understood, however, if one bears in mind that the charge transfer may be due, not only to electrons in the conduction band, but also to ions. It appears that positive free charges must have been predominant in the substances studied by Cohen.

We have previously discussed the contact between two solids with different contact potentials. We have seen that charges of opposite sign arise at their surfaces (in the case of metals) or, in layers near the surface (in the case of semiconductors) and equalize the contact potential difference. It is very probable that the same charges also give rise to the opposite electrification of these solids after their separation.

The direct relation between contact potential and electrification in the case of metals has only been established as recently as 1951. The electrification of semiconductors and insulators has not yet been studied from this most important point of view. Even to this day assertions are made that electrification is due, not to contact, but to friction, and

attempts are made to find a relation between the work of friction and the energy of electrification. Helmholtz gave a widely known explanation of frictional electrification in terms of the separation of a double layer on the surface of the solid, whereby the positive part of the layer is supposed to adhere to one of the rubbing solids and the negative part to the other. An analogous explanation has been given recently by Darmois. He explains electrification in terms of the mechanical severance of valence bonds and the resulting liberation of some of the charges participating in these bonds.

The dependence of the electrification upon the state of the surface, which has been clearly demonstrated by Gezekhus, makes it impossible to arrange materials in a series such that each successive one would acquire negative charge when rubbed on the foregoing one (tribo-electric series), as has been done for contact potentials (Voltaic series) and for thermoelectric e.m.f.s. (Seebeck series).

Tribo-electric series given by different authors are mutually inconsistent. They include various materials which cannot be rigorously controlled, such as cat's fur, silk, glasses, etc. It would be necessary to establish a number of correlations between triboelectric series and Voltaic series. Bearing in mind the fact that charge accumulation may be due to electrons and holes, one could anticipate a further complication in the establishment of such a correlation, owing to the fact that the contact potential difference gives rise in some cases to a layer depleted of free carriers and in other cases to a layer of increased conductivity.

All these problems could easily be studied for semiconductors in which the mechanism of electrical phenomena is well understood.

Semiconductors in electric and magnetic fields[38-41]

We have seen already that the simultaneous presence of an electric field $\mathscr{E}$, giving rise to an electric current J, and of a magnetic field H perpendicular to it, causes the appearance of an electric field $\mathscr{E}_H$ which is perpendicular to both J and H and is given, according to equation (12), by:

$$\mathscr{E}_H = RJH.$$

This expression is valid under the following assumptions: that the temperature of the whole semiconductor remains constant, i.e. the experiment is carried out under isothermal conditions and that, moreover, we are dealing with magnetic fields which are sufficiently weak to enable us to neglect those effects which are proportional to H^2.

Since the thermal velocity of an electron changes by a negligible amount in one free path length, it may be assumed that the electron traverses this path with a constant velocity. The magnetic field transforms the straight path between two collisions into the segment of a

circle. The angle, through which the current is deflected with respect to the electric field is the larger, the longer the time τ between collisions and the lower the velocity of the electron during its flight.

If all conduction electrons have the same velocities, as is the case in metals, then each of them is acted upon by a force F equal to $(e/c)\,Hv$ as well as by the force $\mathscr{E}e$. In a time τ this leads to deflection through an angle

$$\alpha = \frac{e\tau H}{cm^*}. \tag{154}$$

In fact, the force F gives rise to an acceleration dv_H/dt at right angles to the magnetic field and the velocity v:

$$\frac{dv_H}{dt} = \frac{1}{m^*}\cdot\frac{e}{c}\cdot vH.$$

At the same time, the electric field causes an acceleration:

$$\frac{dv_{\mathscr{E}}}{dt} = \frac{e\mathscr{E}}{m^*}.$$

The increment of the electric current J, due to the field H acting for a time dt on a concentration of n electrons in a cubic centimetre, is:

$$dJ_H = JH\frac{e}{m^*c}dt,$$

while the electric field $\mathscr{E}$ causes an increment of current

$$dJ_{\mathscr{E}} = endv = \frac{e^2n}{m^*}\mathscr{E}\,dt.$$

In the steady state, however, the increment of current $dJ_{\mathscr{E}}$ is counterbalanced by an equal and opposite change due to scattering which is determined by the relaxation time τ,

$$dJ_{\mathscr{E}} = -J\frac{dt}{\tau}.$$

Thus, the ratio $dJ_H/dJ_{\mathscr{E}}$, which is equal to the tangent of the angle of deflection of the current (and, at small angles, to the angle itself), is determined by equation (154),

$$\tan\alpha \fallingdotseq \alpha = \frac{dJ_H}{dJ_{\mathscr{E}}} = \frac{e}{m^*c}H\tau.$$

The ratio of the transverse electric field $\mathscr{E}_H$ to the field $\mathscr{E}$ causing the current J is also equal to $\tan\alpha$. The field $\mathscr{E}$ may be expressed in terms of J:

$$J = \frac{ne^2\tau}{m^*}\mathscr{E}$$

and hence:

$$\mathscr{E}_H = \mathscr{E} \tan \alpha = J \frac{m^*}{ne^2\tau} \cdot \frac{e}{m^* c} H\tau = \frac{JH}{nec}. \tag{155}$$

Equating relations (12) and (155) we get:

$$R = \frac{1}{nec}. \tag{156}$$

This simple result has been made possible by the fact that, in a strongly degenerate conductor, all the electrons may be regarded as having the same velocity v and the same relaxation time τ. In the case of semiconductors, however, electrons of all velocities take part in the conduction process, and the value of the Hall constant is affected by the velocity dependence of the relaxation time τ or the free path length l.

In the case of atomic lattices, with only an acoustic branch of the thermal vibrations, the free path length l is independent of the kinetic energy of the electrons. Calculation shows that, in this case,

$$R = \frac{3\pi}{8} \cdot \frac{1}{nec} = \frac{1{\cdot}18}{nec}. \tag{156a}$$

In ionic lattices with optical vibrations, the free path length l is proportional to the velocity v of the electrons. The Hall constant, at not too high a magnetic field and at low temperatures (considerably below the Debye temperature), is obtained in the same form as for metals, in agreement with equation (156):

$$R = \frac{1}{nec}.$$

At high temperatures, on the other hand, where $kT \gg h\nu_0$, $l \propto v^2$ and the Hall constant becomes only slightly less than in atomic lattices.

$$R = \frac{15}{16} \cdot \frac{3\pi}{8} \cdot \frac{1}{nec} = \frac{1{\cdot}10}{nec}. \tag{156b}$$

If the scattering is due wholly to ionized impurities, $l \propto v^4$ and the relaxation time τ is proportional to v^3; the Hall constant becomes:

$$R = \frac{1{\cdot}93}{nec}. \tag{156c}$$

In the case of scattering of the current carriers simultaneously by thermal vibrations and impurity ions, the Hall constant R becomes a complicated function (Fig. 103). The value of this function tends to the limiting value, given by equation (156a) and (156b), as the contribution of scattering by impurity ions decreases and by thermal fluctuations increases. The density of ionized impurities, equal to the density of

free electrons, increases extremely rapidly with temperature as $\exp(-\Delta E/2kT)$, and the contribution of scattering by ionized impurities increases at the same rate. The contribution of scattering by thermal fluctuations, on the other hand, increases as $T^{\frac{1}{2}}$ to $T^{\frac{3}{2}}$, according to the type of lattice, i.e. much less rapidly. The situation is different at sufficiently high temperatures or with a sufficiently low activation energy for the impurities, when the latter become fully ionized. In this case, further increase of temperature gives rise to the more rapid increase of thermal scattering than of ionized impurity scattering.

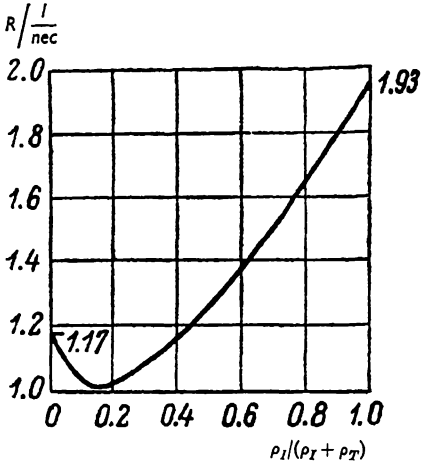


Fig. 103. The dependence of $R/(1/nec)$ upon the ratio $\rho_I/(\rho_I+\rho_T)$, where R is the Hall constant.

The sign of the Hall effect, i.e. the direction of the field $\mathscr{E}_H$, is determined by the sign of the current carriers. The direction, in which a magnetic field H deflects a carrier moving with a velocity v, is reversed by the reversal of the velocity v and of the sign of the carrier itself. With constant directions for $\mathscr{E}$ and H, the particle currents of holes and electrons are moving in opposite directions. In this case, both the velocity v and the charge e change sign simultaneously, so that the direction of deflection of both holes and electrons is the same; the electric field $\mathscr{E}_H$ caused by this deflection has opposite signs in the two cases. By convention the Hall constant for holes has a positive sign, while that for electrons has a negative sign. In this manner the sign of the current carriers may be determined by the direction of the Hall field $\mathscr{E}_H$.

At the same time, one of the formulae (156a) to (156c) enables the determination of the carrier concentration n from the measured value of R.

The measurement of the conductivity σ of the specimen for which R has been determined leads to the value of mobility. In fact,

$$\sigma = neu,$$

$$R\sigma = \frac{A}{nec}.neu = \frac{Au}{c}, \tag{157}$$

where A is the appropriate numerical coefficient from equations (156a) to (156c). It may be noted that none of the expressions for R involves the effective mass m^*.

All the formulae quoted above assume that only one type of carrier takes part in the process of conduction. In some semiconductors, however, the conduction is due to simultaneous motion of holes, the density and mobility of which will be denoted by n_+ and u_+ respectively, and of electrons, for which the corresponding quantities are n_- and u_-. The fields $\mathscr{E}_{H+}$ and $\mathscr{E}_{H-}$ will be mutually opposed and the resulting field may be described by the following expression due to Pisarenko:

$$\mathscr{E}_H = R_0 JH$$

where

$$R_0 = \frac{n_+ u_+^2 - n_- u_-^2}{(n_+ u_+ + n_- u_-)^2} \cdot \frac{A}{ec}. \tag{158}$$

A single measurement of R_0 does not suffice, in this case, for the determination of the sign and of concentration of the current carriers, but it must be combined with the determination of electrical conductivity σ. The latter may be expressed in the form:

$$\sigma = (n_+ u_+ + n_- u_-)e$$

and leads to the formula:

$$R\sigma^2 = \frac{A}{c}(n_+ u_+^2 - n_- u_-^2)e. \tag{159}$$

If other evidence suggests that u_+ equals u_- or, alternatively, if the value of b equal to u_+/u_- is known, then equation (159) enables the difference between n_+ and n_- to be determined. Usually u_+ and u_- are unequal.

The expressions derived above assume different forms in the case of a strong magnetic field. While the deflection of electrons is only slight in weak magnetic fields, in strong fields the electrons suffer, during the time τ of their free flight, deflections through angles of 2π or more.

The calculations of Pisarenko have led to the formula

$$R_0 \fallingdotseq \frac{1}{ec(n_+ - n_-)} \tag{158a}$$

which is valid for metals, and atomic and ionic semiconductors, at both

high and low temperatures. Strong fields, in this context, are fields for which

$$\frac{eH\tau}{m^* c} = \frac{uH}{c} \gg 1. \tag{160}$$

Assuming a mobility about 100 cm²/V sec, it appears that strong fields in this case are such that $H \gg 10^6$ oersted. For a mobility of 10^4 cm²/V sec, $H \gg 10^4$ oersted.

In conclusion, therefore, we may state that the formulae applicable to "weak" magnetic fields may be employed in all cases except at very low temperatures and in extremely strong magnetic fields, approaching 10^6 oersted, such as have been obtained by Kapitza.

The measurement of the Hall constant R represents the simplest and most convenient means of determining the carrier concentration n and the mobility u. The latter frequently plays an important role in technical applications of semiconductors. The differences between the various formulae connecting R and n are, therefore, of practical significance. It is important to know R as a function of n in those cases where the resistivity ρ is due simultaneously to thermal lattice scattering and ionized impurity scattering. Frequently it is permissible, although not strictly correct, to assume that the resistivity ρ is made up of two components: ρ_T due to scattering by thermal vibrations (acoustic vibrations in atomic lattices), and the component ρ_I due to the electric fields of ions. Thus:

$$\rho = \rho_T + \rho_I.$$

Depending upon the magnitude of the ratio $\rho_I/(\rho_T + \rho_I)$, the Hall constant assumes values between

$$R = \frac{1{\cdot}18}{nec} \quad \text{at} \quad \frac{\rho_I}{\rho_T + \rho_I} = 0$$

and

$$R = \frac{1{\cdot}93}{nec} \quad \text{at} \quad \frac{\rho_I}{\rho_T + \rho_I} = 1$$

(Fig. 103).

In many semiconductors R shows a dependence upon the value of the magnetic field. The cause of the phenomenon will be considered in Chapter VII.

A case of particular interest arises when the Hall constant vanishes which, according to equation (158), occurs when:

$$n_+ u_+^2 = n_- u_-^2. \tag{161}$$

Under these circumstances, the magnetic field deflects the paths of electrons and holes in the same direction. Both carrier species reach the sides of the specimen in equal numbers; their charges compensate each other and, thus, do not give rise to an electric field perpendicular

to the current, so that $\mathscr{E}_H$ is zero. The concentration of electrons and holes is higher, however, in that part of the conductor towards which they are deflected and lower at the opposite boundary. The resulting concentration gradient gives rise to diffusion of holes and electrons which equalizes, to a certain extent, the differences between the concentrations. Apart from this, the increased densities of positive holes and negative electrons enhance their recombination, while the probability of their recombination is reduced in the regions in which their densities are lowered. The difference of concentrations, and hence of conductivities, is, therefore, established between the opposite faces of the specimen under the combined influences of the magnetic field, diffusion and of surface and bulk recombination.

If the condition of equation (161) is not strictly fulfilled, a small electric field $\mathscr{E}_H$ will arise, which will accelerate one type of carrier (e.g. electrons) and slow down the other type until they both arrive at the same rate at the surface. The transverse change of resistivity will be less pronounced in this case than in the previous one. Ansel'm has given an analytical treatment of the concentration gradient arising in the presence of a magnetic field.[42]

In our treatment of the influence of magnetic field, we have assumed a uniform temperature throughout the semiconductor. Yet the deflection of the current by the magnetic field gives rise, in turn, to the appearance not only of a concentration gradient but, also, of a temperature gradient directed in the same sense. It is possible, however, to arrange the heat flow from the semiconductor in such a fashion that the temperature difference across it is reduced to a negligible value, so that the process may be considered as an isothermal one.

In the majority of semiconductors the thermal conductivity of the specimen itself is sufficient to safeguard the equalization of temperature. Thus, in most cases, the phenomenon may be regarded as isothermal.[43]

The Hall effect may be used as a convenient means of measuring magnetic field strength. Ge, InSb and HgSe are used for this purpose. The latter compound, in particular, gives highly accurate and reliable measurements. In the Hall effect the electrical energy of the primary current is transformed, with the aid of the magnetic field, into the energy of the current due to the Hall potential difference. In semiconductors such as InSb the energy in the Hall current may reach 20% of that in the primary current.

Change of resistance in a magnetic field

On the basis of the simplest model of electronic motion in a conductor, assuming equal velocities v and equal free path lengths l for all electrons, the resistivity would not be affected by a transverse magnetic field, still

less by a longitudinal one. Acceleration perpendicular to the electric field, under the influence of magnetic field, is fully compensated by the Hall electric field, thus leaving the paths of electrons and the resistivity unchanged. The first measurements of this effect did, in fact, give

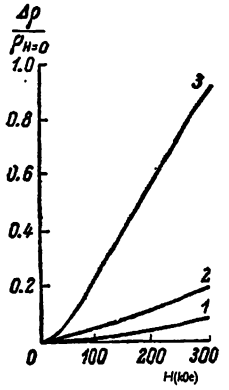


Fig. 104. Relative change of resistivity of cadmium in a transverse magnetic field (according to Kapitza): (1) at 300° K; (2) at 190° K; (3) at 78° K.

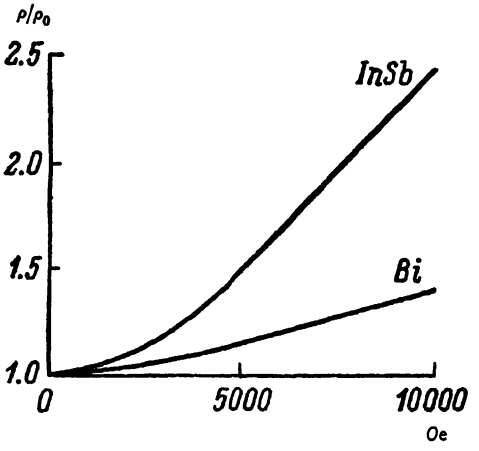


Fig. 105. Change of resistance in a magnetic field for InSb and Bi.

changes of resistivity of the order of only 1%. However, the systematic investigation of a number of metals and some semiconductors, carried out by Kapitza with fields of up to 300,000 oersted, revealed very large effects, increasing with decreasing temperature. Thus, Fig. 104 shows the results obtained by Kapitza for the relative change of resistivity of cadmium in a transverse magnetic field. Fig. 105 shows the field

dependence of the resistivity for Bi and InSb, for which $\Delta\rho/\rho$ is especially large.

The reasons for the effects observed by Kapitza may be understood in the light of a more detailed analysis of the motion of electrons.[44]

The assumption of equal velocities for all conduction electrons in metals is only an approximation, since their velocities are statistically distributed. If holes and electrons contribute to the conduction process simultaneously, their deflections in a magnetic field cannot be compensated, even approximately, by the Hall effect. The Fermi surface, which

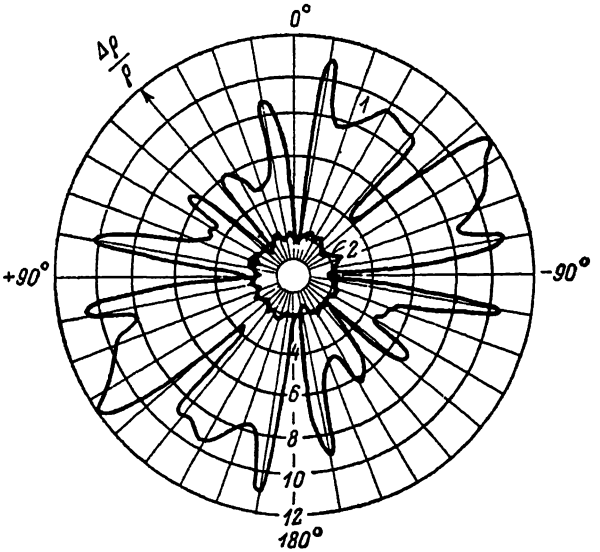


Fig. 106. The relative change of resistivity of a single crystal of gold, cut perpendicular to a principal axis, in a magnetic field oriented along the radius vector. Curve 1 corresponds to liquid helium temperature, 4·2° K, curve 2 to that of liquid hydrogen, 20·4° K.

determines the limiting energy of electrons in momentum space, departs from a spherical shape, even in simple metals belonging to the regular system. The energy of the conduction electrons depends upon direction, this dependence being the stronger, the weaker the thermal motion. At low temperatures this gives rise to a strong dependence of the resistivity upon the crystallographic orientation of the magnetic field. Fig. 106 shows the relation between the relative change $\Delta\rho/\rho$ of the resistivity ρ for a gold single crystal. (The absence of complete symmetry, corresponding to the regular crystallographic system, is due to imperfect orientation of the gold plate with respect to the direction of the magnetic field.)[45]

In semiconductors, apart from the effects of anisotropy, another complication arises as a result of the Maxwellian distribution of electron velocities. If the Hall field compensates for the deflection of electrons of some mean velocity, electrons having lower and higher velocities will suffer deflections from rectilinear paths in one sense or the other. This leads to a reduction of the free path length in the direction of the electric field and, hence, to an increase of the resistance. The effect becomes more pronounced at lower temperatures. At very low temperatures, of the order 1° K, the resistance of some metals (e.g. tungsten) increases tens and hundreds of thousands of times.

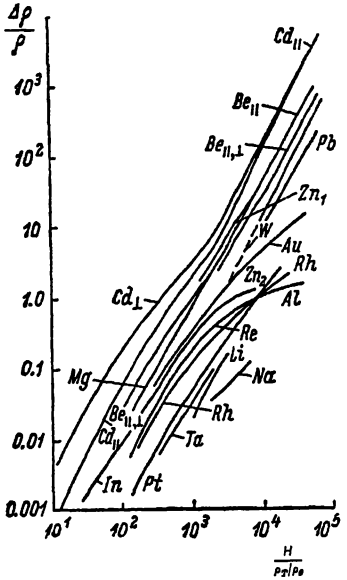


Fig. 107. The effect of a magnetic field upon the resistivity for metals at different temperatures (after Koehler and Justi).

The effect of a magnetic field, upon resistances at different temperatures, is shown in Fig. 107. This gives the relative change of resistivity $\Delta\rho/\rho$ as a function of the ratio $H/(\rho_T/\rho_\theta)$, where H is the magnetic field, ρ_T the resistivity at the temperature T, and ρ_θ the resistivity at the Debye temperature θ. In this coordinate system, to each metal there corresponds one continuous curve, covering the whole range from very low to high temperatures and from weak to relatively strong magnetic fields. Curves for metals of the same type have a similar shape. Since both $H/(\rho_T/\rho_\theta)$ and $\Delta\rho/\rho$ vary by some orders of magnitude, the graph is plotted in logarithmic coordinates as $(\log\Delta\rho-\log\rho)$ against $(\log H-\log\rho_T+\log\rho_\theta)$.

A straight line on this graph corresponds to a power relation between $\Delta\rho/\rho$ and $H/(\rho_T/\rho_\theta)$; the exponent is given by the slope of the straight line. The graph gives a quadratic relation between $\Delta\rho/\rho$ and H for a large number of metals, but some, e.g. Au and Al, show a tendency to saturation at values of $\Delta\rho/\rho$ between 10 and 100.[46]

The currently accepted theory of this effect gives the following dependence of the relative change of resistivity on the magnitude of the transverse magnetic field H, and is valid for semiconductors with an atomic lattice and with two types of current carrier:

$$\frac{\Delta\rho}{\rho}=\frac{9\pi eH^2}{16c^2\sigma}\left[(n_+u_+^3+n_-u_-^3)-\frac{\pi}{4}\frac{(n_+u_+^2-n_-u_-^2)^2}{(n_+u_++n_-u_-)}\right]. \tag{162}$$

For ionic lattices at low temperatures ($kT\ll h\nu_0$) and weak magnetic fields:

$$\frac{\Delta\rho}{\rho}=\frac{n_+n_-(u_++u_-)^2}{(n_+u_++n_-u_-)^2}\cdot\frac{u_+u_-}{c^2}H^2. \tag{163}$$

For $kT\gg h\nu_0$ we have:

$$\frac{\Delta\rho}{\rho}=\frac{27\pi}{6\pi}\cdot\frac{\rho H^2}{c^2\sigma}\left[(n_+u_+^3+n_-u_-^3)-\frac{75\pi}{256}\frac{(n_+u_+^2-u_-u_-^2)^2}{(n_+u_++n_-u_-)}\right] \tag{163a}$$

and for strong magnetic fields $H\gg c/u$:

$$\rho_{H\to\infty}=\frac{32}{9\pi}\cdot\frac{n_+/u_++n_-/u_-}{e(n_+-n_-)^2}. \tag{163b}$$

It will be seen that all the formulae lead to a quadratic relation between $\Delta\rho/\rho$ and H. At the same time Kapitza has found that at fields of the order of tens and hundreds of thousands oersted $\Delta\rho/\rho$ increases linearly with H and that in semiconductors the change of conductivity is very large.

The analysis of the effect in a strong magnetic field, equation (163b), predicts a much smaller variation of resistivity as H tends to infinity. This variation should not exceed 12% with thermal lattice scattering and 70% with ionized impurity scattering, while, in fact, the effect increases linearly far beyond these limits and shows, as a rule, no tendency to saturate.

Equations (162) and (163) refer to semiconductors with both types of current carrier, e.g. intrinsic semiconductors. On the other hand, in the presence of one type of carrier only, whether electrons or holes, the Hall field $\mathscr{E}_H$ opposes the deviations due to the magnetic field H and the change of resistivity is 4·55 times less in the case of atomic lattices and less still for ionic lattices. Thus, for atomic lattices, theory gives:

$$\frac{\Delta\rho}{\rho}=\frac{9\pi}{16}\left(1-\frac{\pi}{4}\right)\left(\frac{uH}{c}\right)^2. \tag{164}$$

The influence of the Hall effect upon the change of resistivity in a magnetic field depends on the geometry of the specimen. It is most pronounced in long thin specimens and least pronounced in plates with current flow in the direction of the thickness. Corbino devised a geometry in which the influence of the Hall effect is altogether absent: it consists of a circular disc with one electrode at the centre and one on the circumference. An electric field in a direction perpendicular to the applied electric field, cannot be set up in view of the complete symmetry of the structure. Fig. 108 gives the dependence of $\Delta\rho/\rho$ upon H in InSb for a number of different geometries.[47]

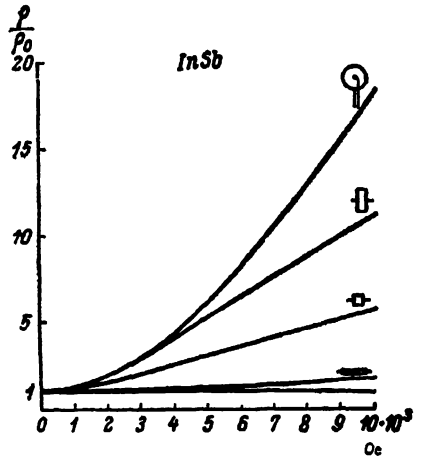


Fig. 108. Change of resistance in a magnetic field for specimens of InSb of various geometries.

According to the simple theory, which assumes spherical surfaces of constant energy, a magnetic field parallel to the direction of current flow would not be expected to give rise to any change of resistivity, in a semiconductor with a lattice belonging to the regular cubic type. Experiment shows, however, that, in general, the resistivity depends strongly upon the magnetic field for all angles between the directions of this field and the primary current. In the case of crystals with strong anisotropy, both theory and experiment lead to the conclusion that their resistivity changes for all directions of the magnetic field. The direction of acceleration of holes and electrons in such crystals is not parallel to the direction of the electric field. Thus, for example, the resistivity of single crystals at low temperatures depends very strongly upon the crystallographic orientation and also varies periodically with the magnetic field strength. Davydov has shown that some of the peculiar properties of bismuth may be accounted for satisfactorily in terms of its energy band structure.

Part of a strongly elongated first band is incompletely filled with electrons, while some electrons are found in a second band. No detailed theory exists, however, which would account for the experimental observations relating to the effect of magnetic field on the properties of bismuth. Single crystals of graphite and of a number of other semiconductors show anisotropy of the mobility and of the effect of a magnetic field in no lesser degree than does bismuth.

Detailed study of the magnetoresistance effects points to the complex shape of the constant energy contours within the Brillouin zone and to the tensor character of the effective mass.

Disagreement with experimental results shows that the existing theory of magnetoresistance effects in semiconductors cannot account properly for the changes in the motion of carriers, due to a magnetic field, unless it takes into consideration the complex nature of the energy surfaces, the dependence of the effective mass on direction and the overlap of the bands.

Apart from the Hall effect and magnetoresistance effect, electronic conductors show a number of other galvanomagnetic and thermomagnetic effects. Thus an electric current I_x normal to a magnetic field H_z gives rise to the appearance of temperature gradients, both in the direction of current flow $(\delta T/\delta x)$—the Nernst effect—and normal to the current flow $(\delta T/\delta y)$—the Ettingshausen effect. Heat flow Q_x along the x-axis, in the presence of magnetic field, causes, in turn, thermal gradients $(\delta T/\delta y)$ and $(\delta T/\delta x)$—Righi–Leduc effects—and potential gradients or electric fields $\mathscr{E}_x$ and $\mathscr{E}_y$—Nernst–Ettingshausen effects. A magnetic field also changes the thermal conductivity κ of an electronic conductor in the case of heat flow parallel to the magnetic field (longitudinal Maggi–Righi–Leduc effect).

All these effects have been studied for various pure metals and alloys but, as regards semiconductors, systematic investigations of the Nernst–Ettingshausen effect have only recently been started (Mochan, Obraztsov and Tsidlikovskij have shown that this effect is especially sensitive to the presence of mixed conductivity, i.e. the presence of carriers of both signs).[48, 49]

All the galvanomagnetic and thermomagnetic effects are closely interconnected. Thus, for example, the sign of the purely thermomagnetic Righi–Leduc effects always agrees with that of the galvanomagnetic Hall effect. The sign of the current carriers may, therefore, be determined from the temperature gradient due to the flow of heat in a magnetic field. We may define the coefficient S of the Righi–Leduc effect through the relation:

$$\frac{\delta T}{\delta y} = SH\frac{\delta T}{\delta x}.$$

S is then related to the Hall constant R by the equation:

$$S = \sigma R$$

where σ is the electrical conductivity.

The thermomagnetic effects in semiconductors are, of course, connected only with that part of the thermal conductivity which is due to the diffusion of electrons.

The Nernst–Ettingshausen effects enable a direct measurement of electronic mobility to be made.

Thus, consider, for instance, the electric field $\mathscr{E}_y$ which appears in the presence of a magnetic field H_z and is due to the heat flow and a thermal gradient $\delta T/\delta x$ normal to both H_z and $\mathscr{E}_y$. In this case

$$\mathscr{E}_y = -NH_z \frac{\delta T}{\delta x}.$$

The constant N may be expressed by the following relation, which is valid for the general case where both holes and electrons are present in semiconductors with atomic lattices, with a constant temperature in the direction of the magnetic field and for low values of H:

$$N = \frac{3\pi}{16} \cdot \frac{k}{c\sigma} \left[-(n_+ u_+^2 + n_- u_-^2) + 8 n_+ n_- u_+ u_- \frac{eu_+ + eu_-}{\sigma} \left(1 + \frac{\Delta E}{4kT}\right)\right].$$

For semiconductors with one type of carrier

$$N = -\frac{3\pi k u}{16 e c} \tag{165}$$

regardless of the sign of current carriers.

In this case the Nernst–Ettingshausen effect enables the mobility of the current carriers to be found.

Galvanomagnetic effects are especially large in semiconductors; their magnitude increases with decreasing carrier concentration and with increasing anisotropy of the crystal lattice. All effects increase with decrease of temperature. Thus, in the neighbourhood of absolute zero, the resistivity of anisotropic bismuth increases by some orders of magnitude and varies markedly with the crystallographic orientation and the strength of the magnetic field. At 1° K the resistivity of tungsten increases in a magnetic field by 10^4–10^5 times.

The Hall effect also reaches high values in semiconductors at low temperatures owing to the increase of the mobility and the decrease of carrier concentration. In contrast to the magnetoresistance effect, the Hall effect in weak magnetic fields depends little upon the anisotropy of the crystal lattice.

Diamagnetism

The problem of the magnetic properties of free electrons has already been considered in the general part of this book. Here we propose merely to deal with the basic facts related to the properties of the crystal lattice.

The physical significance of diamagnetism and paramagnetism has been demonstrated by Langevin. Diamagnetism is the consequence of induction currents arising in the electron shells of each atom. Its magnitude is determined by the Larmor precession. An electron system in a magnetic field H has the precession frequency

$$\omega_L = 2\pi\nu = \frac{eH}{mc}. \tag{166}$$

The diamagnetic susceptibility may be expressed in the following form:

$$\chi = -\frac{Ze^2N}{6mc^2}\overline{r^2} \tag{167}$$

where Z is the number of electrons in an atom or its atomic number in the periodic system, N is the concentration of atoms, $\overline{r^2}$ is the mean square of the radii of individual electronic orbits or, in other words, the mean square of the distance of electrons from the axis of magnetic field.

This formula derived by Langevin on a classical basis agrees with the results of quantum theory. The minus sign is an expression of Lenz's law, according to which the induced magnetic field is everywhere directed in such a sense as to oppose the inducing alternating magnetic field. The electronic cloud frequently screens the nucleus from the magnetic field. The screening by induced currents becomes complete in the superconducting state. The magnetic field does not penetrate at all into a superconductor.

The table below gives experimental values of χ for a number of neutral atoms and ions in a crystal lattice.

He $-1{\cdot}9.10^{-6}$	Li^+ $-0{\cdot}7.10^{-6}$	Mg^{++} $-4{\cdot}3.10^{-6}$	F^- $-9{\cdot}4.10^{-6}$
Ne $-7{\cdot}2$.	Na^+ $-6{\cdot}1$.	Ca^{++} $-10{\cdot}7$.	Cl^- $-24{\cdot}0$.
A $-19{\cdot}4$.	K^+ $-14{\cdot}6$.	Sr^{++} -18 .	Br^- $-34{\cdot}5$.
Kr -28 .	Rb^+ -22 .	Ba^{++} -29 .	I^- $-50{\cdot}6$.
Xe -43 .	Cs^+ $-35{\cdot}1$.		

In order to compute χ it is necessary to know $\overline{r^2}$ and, hence, the distribution of electronic density in a given atom. A satisfactory agreement with experiment is obtained with the use of the Hartree–Fock formula. Thus, the computed value for A is $20{\cdot}6 \times 10^{-6}$ instead of $19{\cdot}4 \times 10^{-6}$, for Li^+ $-0{\cdot}7$, Na^+ $-4{\cdot}9$, K^+ $-15{\cdot}3$, Ce^- $-30{\cdot}4$.

Any change of the electronic charge distribution in a solid leads to a corresponding change of χ. Single and double bonds manifest themselves differently, so that the study of diamagnetic susceptibility of crystals gives some information about the character of the chemical bonding of the lattice. Pascal has formulated a number of rules governing the influence of different crystalline structure factors on the value of χ.

The anisotropy of a crystal lattice leads to a very pronounced anisotropy of the diamagnetic susceptibility which assumes exceptionally high values in certain crystallographic planes. Pronounced anisotropy of χ is found in arsenic, bismuth and, especially, in graphite with its layer lattice. The high value of χ in the latter case has been explained by Ehrenfest in terms of the overlap of electronic orbits which enclose several atoms. Similar anisotropy of χ is typical of the molecules of aromatic compounds.

Paramagnetism

According to Langevin, paramagnetism consists in the presence of a magnetic moment μ, which may be due to electronic spin or to the orbital motion of electrons. As a general rule, however, it may be stated that a magnetic moment is found in atoms and molecules with an odd number of electrons or with incompletely filled inner electronic shells, which indicates the paramount importance of spin. Only a few compounds with an even number of electrons, amongst them oxygen molecules, show the presence of a magnetic moment. The conclusion, regarding the ratio of the orbital magnetic moment M_{orb} to the spin magnetic moment M_{spin}

$$\frac{M_{\text{orb}}}{M_{\text{spin}}} = \epsilon$$

where ϵ is a small quantity, is confirmed by the direct measurement. According to Langevin's theory, which is in agreement with the quantum theory, the paramagnetic susceptibility, in weak magnetic fields satisfying the relation $\mu H \ll kT$, is given by:

$$\chi = \frac{N\mu^2}{3kT}. \tag{168}$$

In contrast to the diamagnetic susceptibility, which is independent of temperature, the paramagnetic susceptibility shows, as a rule, an inverse dependence on the absolute temperature. This law is named after Pierre Curie, who established it experimentally.

The magnetic moment of a Bohr magneton is

$$\mu = \frac{eh}{2mc} = 0{\cdot}927 \times 10^{-20} \text{ erg/gauss}. \tag{169}$$

The magnetic moments of ions are not conveniently expressed in terms of this unit. Thus, for example, the effective magnetic moments μ_{eff} for ions of the iron group have the following values in terms of Bohr magnetons:

	μ_{eff}
Ti^{+++}, V^{++++}	1·8
V^{+++}	2·8
Cr^{+++}, V^{++}	3·3
Mn^{+++}, Cr^{++}	4·9
Fe^{+++}, Mn^{++}	5·9
Fe^{++}	5·4
Co^{++}	4·8
Ni^{++}	3·2
Cu^{++}	1·9

In our discussion of energy levels in crystals, we made the assumption that the electronic energy is independent of the orientation of the spin and, neglecting their interaction, we placed two electrons of opposite spins on each energy level.

In a magnetic field each level splits into two levels of different electron energy. The energy difference ΔE, corresponding to two electrons having the same wave vector. k_x, k_y, k_z but opposite orientations of spin, is

$$\Delta E = 2|\mu_0|H = g\mu H \tag{170}$$

where g is the spectroscopic splitting factor and is equal to 2 for electronic spin (more exactly $g = 2{\cdot}003$).

If an orbital magnetic moment is present at the same time, then g is determined by Landé's formula:

$$g = 1 + \frac{I(I+1) + S(S+1) - L(L+1)}{2I(I+1)} \tag{171}$$

where I, S and L are the total, spin and orbital quantum numbers respectively.

There are discrete optical transitions from levels with spin S to magnetically split-off levels of spin $(S \pm 1)$. Energy is absorbed, therefore, if the electric field in the semiconductor oscillates with a frequency ν such that:

$$h\nu = g\mu H. \tag{170a}$$

This increases considerably the energy losses in the high-frequency circuit which generates the field. This phenomenon, known as paramagnetic resonance, was discovered by Zavojski[50] in 1945.

The apparatus for paramagnetic measurements consists principally of a microwave cavity carrying the specimen and placed between the pole-pieces of a powerful electro-magnet. The magnetic field of the cavity is normal to that of the electro-magnet.

The resonant frequency in a field of 10^4 oersted is

$$\nu_r = \frac{2 \times 0{\cdot}927 \times 10^{-20} \times 10^4}{6{\cdot}61 \times 10^{-27}} = 2{\cdot}8 \times 10^{10} \text{ c/s.}$$

Thus paramagnetic resonance may be found at frequencies in the centimetre waveband. Fig. 109a shows the dependence of the magnetic losses in an organic molecule as a function of the magnetic field H at a fixed microwave frequency $2{\cdot}4 \times 10^{10}$ c/s. The experimental curve shows a sharp peak with a half-amplitude width of 42 Oe, i.e. 0·5% of the magnetic field strength at the maximum, 8600 Oe.

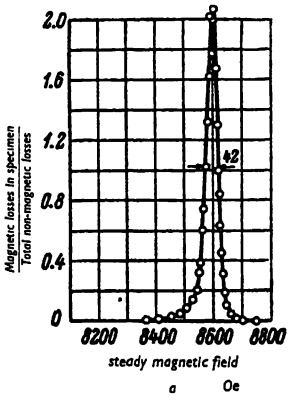


Fig. 109a. Dependence of magnetic losses on magnetic field strength.

Apart from spin resonance, there exists, also, resonance associated with the magnetic moment of the nucleus, which is some three orders of magnitude weaker than the spin magnetic moment. According to equation (168), the contribution of the nuclear moment to the paramagnetic susceptibility is about 10^6 times less than the contribution of the spin moment. Nevertheless Lazarev and Shubinkov have succeeded in observing, in solid hydrogen, paramagnetism as well as the diamagnetism of the electron cloud.

The nuclear magnetic moments of a number of nuclei are given in Table 10.

Superimposed on the spin magnetic resonance is the nuclear magnetic resonance which is measured in units of nuclear magnetons

$$\mu_n = e\hbar/2Mc = 5{\cdot}05 \times 10^{-24}. \tag{172}$$

One nuclear magneton corresponds to a frequency ν_n given by

$$h\nu_n = \mu_n H. \tag{172a}$$

The frequencies associated with nuclear magnetic resonance are about a thousand times lower than those of spin resonance and fall in the short-wave radio-frequency band of the order of 10 Mc/s.

The line width for spin resonance is determined by the magnetic fields of neighbouring electrons superimposed on the steady magnetic field. If a is the distance between neighbouring atoms, their magnetic fields amount to approximately μ/a^3. With μ equal to 10^{-20} and a to 3×10^{-8} cm, the fields of neighbouring atoms contribute an additional field strength of approximately 400 Oe, which determines the average width of the absorption peak.

The influence of neighbouring atoms on nuclear resonance is weaker by a factor 10^3, i.e. it is of the order of 1 Oe.

TABLE 10

Nucleus	Spin in units	Magnetic moment in nuclear magnetons
Neutron	$\frac{1}{2}$	−1·913
Proton	$\frac{1}{2}$	+2·793
Deuteron	1	0·857
Li^7	$\frac{3}{2}$	3·256
Na^{23}	$\frac{3}{2}$	2·217
Mn^{55}	$\frac{5}{2}$	3·468
Co^{59}	$\frac{7}{2}$	4·648

In the case of liquids, the vibrations of molecules occurring at frequencies 10^{12}–10^{13} sec^{-1} cancel out, on the average, the contributions of neighbouring molecules to the magnetic field. The modulation of the magnetic field, which is caused by these vibrations, has too high a frequency to influence the resonance. For this reason, the resonance curves of liquids also show sharper peaks.

Local imperfections in a crystal lattice, impurities and free charges cause additional resonance maxima or give rise to the broadening of the existing main absorption peak. This has led to the application of paramagnetic resonance as a basic method for the detection of all kinds of impurities in the semiconductor crystal lattices. Nuclear resonance, which lies in an easily accessible part of the spectrum, is especially well suited for this purpose.

The crystal lattice surrounding a given atom produces a magnetic field in it, even in the absence of an external electro-magnet. This magnetic field, too, leads to paramagnetic resonance which has to be taken into account in the interpretation of resonance spectra.

Fig. 109b gives the paramagnetic spin resonance spectrum of ammonium–chromium alum at a frequency $2{\cdot}4 \times 10^{10}$ c/s. The spectrum shows fine structure, due to the spins of neighbouring atoms, superimposed on the resonant frequency.

The resonance spectrum of the nuclear magnetic moment is shown in Fig. 110. In contrast with the case of spin resonance, the axis of abscissae here corresponds to frequency, the magnetic field H being kept constant. Nitrated iron in a magnetic field 7050 Oe showed a resonance maximum at a frequency of 30 Mc/s.

Frequently it is possible to obtain interaction of spin paramagnetic resonance with the nuclear moments. This interaction appears as a hyperfine structure of the resonance spectrum. Fig. 111 shows the hyperfine structure of the resonance maximum of Mn^{++} ions in water.

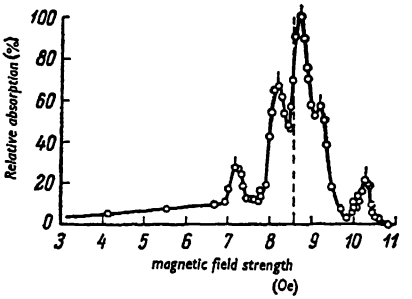


Fig. 109b. Absorption in chromium–ammonium alum $NH_4Cr(SO_4)_2$ 12 H_2O at wavelengths of the order 1 cm.

Cyclotron or diamagnetic resonance[51,52]

Cyclotron resonance is determined by the precession of the electronic magnetic moment in an external magnetic field H. The resonant frequency ν of the magnetic field acting on the electron is given by

$$\nu = 2\pi \frac{e}{m^* c} H$$

(assuming that the ratio of magnetic moment to electron spin corresponds to $g = 2$).

An electronic charge e, orbiting around the lines of force of the same magnetic field H, also experiences resonance if the alternating electric field increases as the frequency ν is approached, as in the case of paramagnetic resonance.

In paramagnetic resonance the high-frequency magnetic field should be perpendicular to the external field H. In cyclotron resonance, on the

other hand, the high-frequency electric field should be normal to the field H.

In contrast to the case of paramagnetic resonance, only free electrons and holes participate in cyclotron resonance and, even then, only if they are able to describe at least one full circle, or a substantial part of it, without collisions.

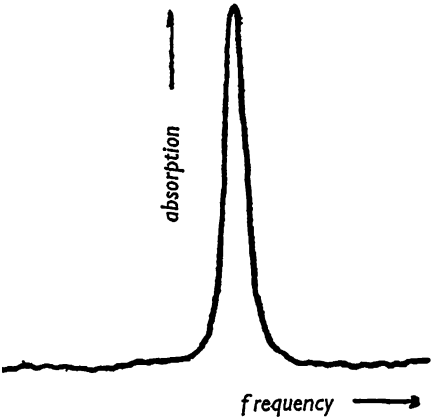


Fig. 110. Resonance spectrum for the nuclear magnetic moment.

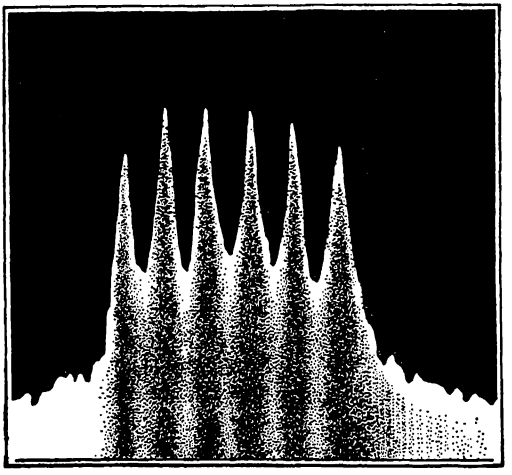


Fig. 111. Hyperfine structure of the resonance maximum of Mn^{2+} ions in water. The nuclear spin $I = \frac{5}{2}$, electronic spin $S = \frac{5}{2}$.

With an electron velocity v cm/sec and a frequency of rotation ν sec^{-1}, the circumference of the circular orbit is v/ν. The condition for the existence of cyclotron resonance is, thus, $l > v/\nu$.

Assuming m^* equals m_0 and H is 10^4 Oe, we obtain for the cyclotron frequency ν an order of magnitude 3×10^{10} c/s. If v is about 10^7 cm/sec, the free path length must satisfy the condition $l > 3 \times 10^{-4}$ cm.

In terms of the mobility u, instead of the free path length l, the condition for the appearance of resonance is $u > 10^4$ cm^2/V sec, while at low temperatures, for which v is about 10^6 cm/sec, it is necessary for $u > 10^5$ cm^2/V sec. For arbitrary values of the effective mass m^*, the former conditions become:

$$l > 3 \times 10^{-4} \frac{m_0}{m^*} \text{ cm}, \quad u > 10^4 \frac{m_0}{m^*} \text{ cm}^2/\text{V sec}.$$

The ratio m_0/m^* is greater than 1 for all semiconductors with high mobilities. On this account, these conditions are not satisfied for any of the known semiconductors at room temperature. Resonance can be obtained, however, at low temperatures where l increases and v decreases sufficiently.

Another difficulty arises, however, at these low temperatures. The concentration of electrons becomes very low and does not lead to sufficient absorption of energy. This difficulty is surmounted by supplying excess carriers to the specimen by means of illumination. In semiconductors with low impurity activation energies, such as germanium, even the high frequency vibrations themselves are sufficient for the partial excitation of the impurities and the increase of the free carrier concentration. At a frequency ν of 3×10^{10} c/sec, the magnitude of $h\nu$, $1{\cdot}2 \times 10^{-4}$ eV, becomes comparable with the ionization energy.

Cyclotron resonance of free electrons and holes has been obtained, in fact, at liquid helium temperatures, at and below 4° K, in germanium, silicon and indium antimonide.

Electrons and holes rotate in the opposite sense in a magnetic field. Their cyclotron resonances may be observed separately if circularly polarized high-frequency radiation is used instead of linearly polarized waves.

Cyclotron resonance offers the most direct method of studying the magnitude and character of the effective mass of carriers. The line width of the cyclotron resonance absorption peak is determined by the number of free revolutions completed by the carrier and this leads to an estimate of its free path. If τ denotes the time of free light of a carrier, the half-amplitude width of the resonance curve is given by:

$$\Delta\nu = 1/\tau.$$

At room temperature, the order of magnitude of τ is given by:

$$\tau = l/v \fallingdotseq 10^{-6}/10^7 = 10^{-13} \text{ sec},$$

whence $\Delta\nu$ is about 10^{13}, i.e. higher than ν which is 3×10^{10}. At liquid helium temperatures, on the other hand,

$$\tau \fallingdotseq 10^{-4}/10^{6} = 10^{-10} \text{ sec},$$

so that $\Delta\nu/\nu$ is about $\frac{1}{3}$.

The results of the measurements of effective mass by the cyclotron resonance method show good agreement with data obtained by both paramagnetic resonance and magnetoresistance methods.

The study of cyclotron resonance has revealed the complex nature of energy contours in the Brillouin zone. This has led to a complication of the concept of the effective mass, the latter being determined by the second derivative of the energy E with respect to the wave vector k.

The conduction band of a relatively simple crystal such as germanium, which has a cubic diamond-type lattice, consists of 8 ellipsoids with axes directed along the cube diagonals. In each of these ellipsoids, the effective mass represents a tensor, i.e. the relation between E and k has the form:

$$E(k) = \frac{h^2}{2}\left(\frac{k_1^2}{m_\parallel^*} + \frac{k_2^2}{m_\perp^*} + \frac{k_3^2}{m_\perp^*}\right),$$

where k_1 is the component of the k-vector along the cube diagonal [111] and k_2 and k_3 are the components in the plane perpendicular to the [111] direction. For electrons, the longitudinal mass $m_\parallel^*$ is found to be $1{\cdot}57\ m_0$, while the transverse mass $m_\perp^*$ is $0{\cdot}08\ m_0$.

The effective mass, determined from the resonant frequency, depends upon the orientation of the magnetic field and is given by:

$$\frac{1}{(m^*)^2} = \frac{\cos^2\phi}{(m_\parallel^*)^2} + \frac{\sin^2\phi}{m_\parallel^* m_\perp^*}$$

where ϕ is the angle between the magnetic field and the [111] axis.

The relationships for holes near the top of the valence and the corresponding shapes of the energy contours are more complicated. The relation between the energy and the components of the wave vector is then:

$$E(k) = Ak^2 \pm [B^2k^4 + C^2(k_x^2 k_y^2 + k_y^2 k_z^2 + k_z^2 k_x^2)]^{\frac{1}{2}}.$$

Two types of hole have values of the effective mass m^* equal to $0{\cdot}04\ m_0$ and $0{\cdot}3\ m_0$ respectively. The effective masses are almost isotropic in character.

The effective masses for electrons in the conduction band of silicon are:

$$\text{longitudinal } m_\parallel^* = 0{\cdot}98\ m_0,$$

$$\text{transverse } m_\perp^* = 0{\cdot}19\ m_0.$$

The axis in this case is, however, in the [100] and not the [111] direction.

In the valence band, m^* is about 0·16 m_0 for light holes and 0·5 m_0 for heavy holes.

The effective mass for electrons in InSb has the value 0·013 m_0 and is independent of direction; for holes two values have been obtained; these are 0·18 m_0 and 1·2 m_0 respectively.

Ferromagnetism

Ferromagnetics are materials in which large groups of atomic magnetic moments are co-ordinated by bulk forces and form regions (domains) in which there is orientation of their magnetization. In the absence of a magnetic field, the directions of magnetization of the domains are distributed randomly.

The application of a magnetic field gives rise to orientation of the direction of magnetization in large groups of atoms. The magnetic permeability of ferromagnetics, therefore, exceeds by millions of times the values obtained with paramagnetic materials. These domains with their orientation in the direction of the magnetic field predominate, even after its removal, and give rise to the phenomenon of residual magnetization or coercive force.

The presence of domains and their geometry and dimensions may be revealed directly by the application of a drop of a suspension of ferromagnetic particles to the surface of a ferromagnetic. The particles collect along the domain boundaries, where the magnetic field and its gradient have their maximum values, thus bringing out the pattern. Their dimensions are of the order of tens of microns and may be predicted theoretically on the basis of the principle of minimum magnetic and surface energy. The boundaries of domains are displaced and grow under the action of the magnetic field, giving rise to the magnetization of the whole specimen.

Saturation sets in when all domains are orientated in the direction of the magnetic field. The magnetic field necessary for saturation decreases with decreasing temperature, owing to the reduction of thermal motion which disturbs the orientation of domains.

A definite relation exists between the coercive force H_c of a given material and its magnetic permeability in weak magnetic fields. The higher the coercive force H_c, the lower the permeability μ.

Each ferromagnetic material has a definite Curie temperature above which the material is paramagnetic. The paramagnetic susceptibility χ above the Curie point decreases with temperature, as in all paramagnetic materials. The difference consists in the fact, however, that the susceptibility of paramagnetic materials follows the Curie law

$$\chi_{\text{par}} \propto \frac{1}{T} \tag{173}$$

(except at very low temperatures), while in ferromagnetics, above the Curie point, the susceptibility is given by

$$\chi_{\text{fer}} \propto \frac{1}{T - T_k} \tag{174}$$

where T_k is close to the Curie temperature.

If the exchange integral, which determines the ferromagnetic properties of a material, has a negative sign, the state of minimum energy corresponds to anti-parallel orientation of magnetic moments. In this case the domains do not consist of groups of identically orientated spins but of oppositely orientated pairs which give zero resultant magnetic moment. Such materials are known as "anti-ferromagnetics". They, too, have a critical Curie temperature above which the regular arrangement of spins vanishes and the material becomes paramagnetic. The magnetic susceptibility of anti-ferromagnetics obeys the law

$$\chi_{\text{anti-fer}} \propto \frac{1}{T - \Theta}. \tag{175}$$

In contrast to ferromagnetics Θ does not equal T_k.

The temperature-dependence of the paramagnetic susceptibility for paramagnetics, ferromagnetics and antiferromagnetics is shown in Fig. 112.

The magnetic susceptibility of an antiferromagnetic material shows a sharp maximum at the Curie temperature T_k. Fig. 113 shows the function $\chi(T)$ for manganese selenide. Fig. 114 shows the orientation of spins of the Mn^{++} ions in the crystal cell of the compound MnO.

The magnetic moments of the atoms of ferromagnetic materials may be due to two causes: electronic spin and orbital motion. Comparison between the magnetic moment and the mechanical moment of rotation enables an estimate to be made of the relative contributions to the total magnetic permeability. The ratio of the magnetic moment M to the angular momentum I for a circular orbit is:

$$\frac{M_{\text{orb}}}{I_{\text{orb}}} = \frac{e}{2mc}. \tag{176}$$

For spin, on the other hand:

$$\frac{M_{\text{sp}}}{I_{\text{sp}}} = 2\frac{e}{2mc} = \frac{e}{mc}. \tag{177}$$

Both formulae may be expressed as

$$\frac{M}{I} = \frac{ge}{2mc},$$

where $g = 2$ for spin and $g = 1$ for orbital motion.

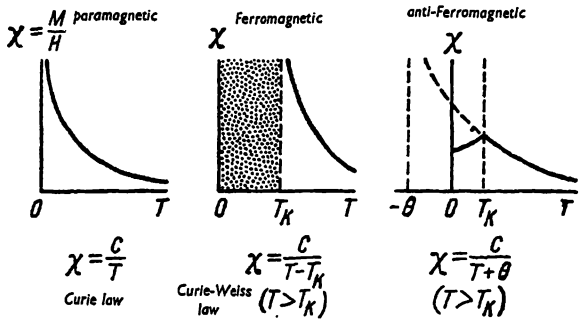


Fig. 112. Temperature dependence of the magnetic susceptibility of paramagnetic, ferromagnetic and anti-ferromagnetic substances.

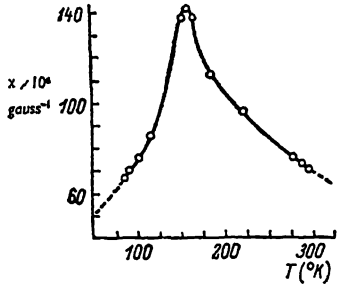


Fig. 113. The temperature dependence of the magnetic susceptibility of manganese selenide.

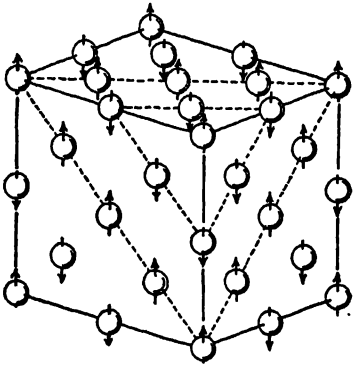


Fig. 114. Disposition of spins of Mn^{++} ions in manganese oxide MnO, determined by neutron diffraction.

Having determined experimentally the ratio M/I, we may deduce which of the two mechanisms is responsible for the magnetization of the material or to what extent both of them contribute.

These measurements may be carried out according to two basic methods. The method of Einstein and de Haas determines the angular momentum of a rod on reversing the sign of its magnetization, while Barnett's method measures the magnetization of a ferromagnetic rod during rotary motion. The results from both methods are in good agreement. The value of the coefficient g, multiplying the gyroscopic constant, lies between the limits 1·85 and 2·00, which indicates the predominance of spin magnetism over orbital magnetism. The relative contribution of the orbital motion to the magnetization will be denoted by

$$\frac{M_{\text{orb}}}{M_{\text{sp}}} = \epsilon.$$

In the case of metals and semiconductors one might also expect to find other values of g and even a reversal of sign. We have considered electrons with a negative charge $-e$ and holes with a positive charge $+e$. One would expect the gyromagnetic effect to have the opposite sign in those semiconductors and metals conducting electricity by means of holes and in those which conduct by means of electrons.

Moreover, we have introduced the concept of an effective mass m^*, which in some cases is several times less than m_0 and, in other cases, is larger than m_0.

Both propositions are decisively disproved by experiment. Neither the sign of the current carriers, nor their effective mass, has any influence on the quantity $e/2mc$, as determined by one of the indicated methods from the gyroscopic effect. This experimental result proves convincingly that the concepts of holes and of effective mass do not characterize the mechanical properties of real electrons but determine the interaction between electrons and the periodic field of the lattice. These concepts may only be applied to the determination of the mean velocity and mean acceleration of electrons in a semiconductor under the influence of electromagnetic forces.

Another method of determining the gyroscopic ratio g employs ferromagnetic resonance which is entirely similar to paramagnetic resonance. In the former, as in the latter,

$$\omega = g\frac{e}{2mc}H.$$

It has been shown by Kittel that, taking into account the de-magnetization effects occurring in the discs which are used in experiments with

ferromagnetic materials,

$$\omega = g\frac{e}{2mc}H\sqrt{\left(1+4\pi\frac{M}{H}\right)}$$

where M is the magnetic moment of the ferromagnetic in a field H.

Table 11 gives the values of the gyroscopic ratio and the fraction ϵ of orbital magnetism for various ferromagnetics, obtained both by gyroscopic experiments and by resonance.

TABLE 11

Material	g		ϵ	
	gyroscopic experiments	ferromagnetic resonance	gyroscopic experiments	ferromagnetic resonance
Fe . . .	1·93	2·12	0·04	0·06
Co . . .	1·87	2·22	0·07	0·11
Ni . . .	1·92	2·2	0·04	0·1
Fe_3O_4 . .	1·93	2·2	0·04	0·1
Cu_2AlMn . .	2·00	2·01	0·00	0·005
Permalloy . .	1·91	2·07	0·05	0·04

It is evident that systematic experimental errors could not be completely eliminated to the resonance measurements.

Ferrites

There exists a class of semiconducting materials which have ferromagnetic or ferrimagnetic properties. Unlike ferromagnetic metals, they have a high resistivity, exceeding that of iron by at least 10^6–10^7 times. They differ, also, in their mechanism of electrical conduction and in a number of other specific properties. These materials are known as ferrites.[53]

Ferrites have the crystal lattice of the mineral spinel ($MgAl_2O_4$). The unit cell of cubic symmetry consists of 32 close-packed divalent oxygen ions with a cube edge length 8·4 Å. The spaces between the oxygen ions allow room for 16 metal ions with four neighbours (tetragonal disposition) and 32 metal ions with six neighbours (octahedral disposition). Sixteen trivalent and 8 divalent metal ions, placed in half of these empty spaces, compensate for the 64 negative charges of the oxygen ions.

We distinguish between simple and inverse spinels. Ferrites belong to the latter type: here 8 divalent and 8 trivalent ions are distributed in the octahedral spaces, while the 8 remaining trivalent ions occupy the tetrahedral spaces.

The chemical formula of ferrites may be represented in the form $M^{++}M_2^{+++}O_4^{--}$, where M^{++} is a divalent, and M^{+++} a trivalent, metal ion. Fe_3O_4 is a special case of a ferrite which may be represented as $FeOFe_2O_3$, or $Fe^{++}Fe_2^{+++}O_4^{--}$. Fe^{++}, in a ferrite, may be replaced by Mg^{++}, Zn^{++}, Cd^{++}, Mn^{++}, Co^{++}, Ni^{++}, Cu^{++} or a mixture of these ions. Ferrites with mixtures of Zn^{++} and Mn^{++} and of $Zn^{++}Ni^{++}$ have been found. Smolenskij has studied a large number of different ferrites.

The simplest ferrite, Fe_3O_4, has a saturation magnetization of 485 Oe, corresponding to 4·2 magnetons per molecule. However, one Fe^{++} and two Fe^{+++} ions, with parallel orientation, should exhibit 14 magnetons. This discrepancy is explained by the fact that the trivalent Fe^{+++} ions are in anti-parallel positions and their moments cancel each other; thus, the magnetic moment of the molecule is determined by one Fe^{++} ion with 4 magnetons; this is the situation known as ferrimagnetism.

Table 12 gives values of the saturation magnetization M_s, the effective number of magnetons and the Curie temperature of some ferromagnetic metals and of a number of ferrites.

TABLE 12

Material	Maximum magnetization	Number of magnetons	Curie temperature (°K)
Fe	1750	2·221	1043
Co	1450	1·716	1400
Ni	510	0·606	631
Gd	1980	7·1	289
Cu_2MnAl	580	4·0	603
MnBi	675	3·52	630
MnSb	710	3·53	587
MnAs	870	3·40	318
Fe_3O_4	485	4·2	848
$MnOFe_2O_3$	358	5·0	783
$NiOFe_2O_3$	240	2·3	863
$CuOFe_2O_3$	290	1·3	728
$MgOFe_2O_3$	143	1·1	583

Thermal conductivity of semiconductors

The thermal parameters of semiconductors are as important as the electrical ones in some applications. Thus, the efficiency of a thermoelement depends on the thermal e.m.f., on the electrical conductivity, and on the thermal conductivity. The same is true for thermistors, for time-switches, etc. Thermal conductivity is also important in rectifiers.

It so happens that relatively little has been done either theoretically or experimentally on thermal conductivity. The published results relate mainly to the variation in thermal conductivity with temperature, and

even here most of them relate to very low temperatures. Aspects such as the variation with composition or structure have scarcely been touched on; no attempt has been made to predict a thermal conductivity. I give here a few of my own ideas, which have not yet been completely verified, together with the known facts and theoretical deductions.

The temperature T is given as a function of the coordinates x, y, z and of the time t by the differential equation

$$\frac{\partial T}{\partial t} = \frac{\kappa}{c\rho}\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}\right).$$

Here T is known if the initial and boundary conditions are known; κ denotes the thermal conductivity, c is the specific heat, and ρ is the density. Then $c\rho$ is the thermal capacity of unit volume, and K equal to $\kappa/c\rho$ is called the thermal diffusivity.

Samples in the form of plates or rods are usually used in measurements. Consider the steady-state heat flow in a cylinder of cross-section area S; if we neglect heat losses and assume that an amount of heat Q passes through, we have

$$\frac{dQ}{dt} = \kappa \frac{dT}{dx} S, \tag{178}$$

where κ as before is the thermal conductivity, while x is the direction of the heat flow; dT/dx is the temperature gradient.

Two essentially distinct conduction mechanisms operate in a semiconductor, namely (1) conduction via longitudinal and transverse elastic waves, and (2) conduction via electron diffusion. We may assume that κ is made up additively from κ_1, which would be the conductivity in the absence of electron conduction, and κ_2, which is the conductivity due to the electrons or holes:

$$\kappa = \kappa_1 + \kappa_2. \tag{179}$$

Semiconductors whose carriers are all of one sign have κ_2 related to the electrical conductivity σ by the Wiedemann-Franz law

$$\frac{\kappa_2}{\sigma} = A\left(\frac{k}{e}\right)^2 T, \tag{180}$$

where k is Boltzmann's constant, e is the electronic charge, and T is the absolute temperature. A is equal to 2 in a non-degenerate electron system if the mean free path l is independent of the energy ϵ; in general A equals $(r+2)$, where r is the exponent in the law $l \propto \epsilon^r$. Thus, in ionic crystals above the Debye temperature l is proportional to energy so that A equals 3. The degenerate system in a metal gives A equal to $\pi^2/3$.

Measurements confirm these predictions; Maslokovets and Dunaev[54] have measured the thermal conductivity of lead sulphide as

a function of impurity content (i.e. as a function of σ) and have found that

$$\kappa = \kappa_0 + \beta\sigma \tag{179a}$$

where κ_0 is the thermal conductivity corresponding to zero electrical conductivity, while β has a value close to $2(k/e)^2 T$. The empirical law of equation (179a) is naturally to be taken as equivalent to equations (179) and (180), with β equal to $A(k/e)^2 T$. Devyatkova has found a relation

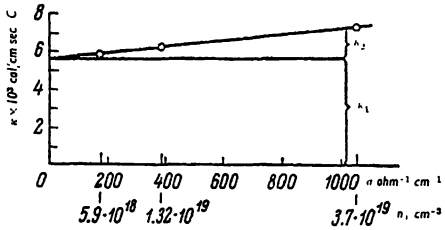


Fig. 115. Relation of thermal conductivity to electrical conductivity and to carrier concentration for p-type PbS.

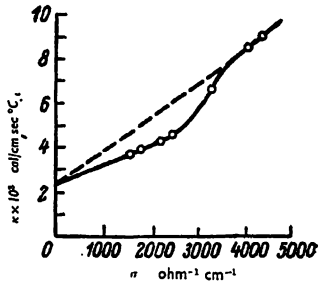


Fig. 116. Relation of thermal conductivity to electrical conductivity for a semiconductor.

corresponding to equation (179a) for lead telluride having an electron concentration that brings the material into the degenerate class. Here β takes nearly twice the value corresponding to non-degeneracy.[55]

A. V. Ioffe and I have measured κ for a series of semiconductors of the same composition with electrical conductivities ranging from 1500 to 4400 $\text{ohm}^{-1}\text{cm}^{-1}$, and have derived the relation between κ and σ shown in Fig. 116, which relation can also be represented as in equation (179a). Here β has a value of $2(k/e)^2 T$ in the non-degenerate range, and a value about 50% larger in the degenerate range; the value changes continuously from $2(k/e)^2 T$ to $3{\cdot}3\,(k/e)^2 T$ in the transition region. The boundary between the two main regions lies at concentrations in the range $3 \times 10^{19} > n > 2 \times 10^{19}\,\text{cm}^{-3}$, as we would expect from the laws of quantum statistics. To a first approximation κ_2 increases with σ (for

$n < 2 \times 10^{19}$ cm^{-3}) at the rate of 0·001 cal/cm sec °C per 1000 ohm^{-1} cm^{-1}, and at the rate of 0·0016 cal/cm sec °C for $n > 3 \times 10^{19}$ cm^{-3} (in the degenerate region).[56]

The data are few and not very accurate, but we may say that quantum theory correctly predicts that part of κ which is dependent on the electrical conductivity. That part increases with temperature as a consequence of the rise in σ and as a consequence of the factor T in equation (180).

It is much more difficult to estimate κ_1, the lattice conductivity. Debye put forward the basic ideas in 1914 soon after he published his theory of specific heat. In both cases he considered thermal motion as the assembly of all possible inherent oscillations in the body. Each such oscillation is equivalent to an elastic wave propagating in the crystal. In this mode of conduction the heat is carried by elastic waves from the hot parts (where the waves are stronger) to the cold parts. However, since these waves travel at the speed of sound, the body would very soon become equally hot everywhere, and the waves would merely be reflected from end to end. This does not in fact happen. It may take hours for an insulator to come to the same temperature at all points. Therefore Debye assumed that the waves are repeatedly scattered on their way from the hot end to the cold one, i.e. that the mean free path is of the same order as for electrons (10^{-7} cm). Debye supposed that the cause of such strong scattering was interaction with waves travelling in the opposite direction. If the waves may be taken as harmonic, the compressions and rarefactions (longitudinal waves) or displacements (transverse waves) have a constant speed, and so they cannot deflect one another. A sound wave, if strictly harmonic, passes freely across any other linear wave system, e.g. across a light beam. The scattering of light by light has not been detected. Peierls has shown that energy cannot be redistributed between the frequencies when harmonic waves interact, and that equilibrium cannot thereby be reached. Therefore thermal oscillations could not come to equilibrium in a medium without a discrete, atomic, structure.

Scattering can occur only if the atoms move anharmonically, and if the displacements are so large as to deviate from the proportionality implied by Hooke's law. Only in such a case can the various parts of a wave have different speeds of propagation, which speeds are determined by the ratio of the elasticity to the density. Density fluctuations exceeding the range to which Hooke's law applies deflect or partly reflect waves. The atoms in a solid in fact change their relative distances in thermal oscillation to a much greater extent than in normal elastic deformation. Therefore it is reasonable to suppose that thermal elastic waves are highly anharmonic, which enables us to explain heat conduction.

No matter what the details of the scattering process, there is undoubtedly a far-reaching analogy between the scattering of these

waves (phonons in Tamm's corpuscular representation) and diffusion or thermal conduction in gases.

Debye has shown that we may treat thermal resistance as the result of multiple scattering of randomly moving phonons at thermal fluctuations, and that we may express κ_1 in terms analogous to those for diffusion or thermal conduction in gases, which take the form

$$\kappa_1 = \tfrac{1}{3} c' v \bar{\lambda}, \tag{181}$$

where v is the speed of a phonon, i.e. the speed of propagation of elastic waves, c' is the thermal capacity per unit volume, and $\bar{\lambda}$ is the mean free path. (The factor 1/3 given by the kinetic theory of gases is replaced by 1/4 in Debye's representation). Debye assumed that the waves of shortest wavelength travel at the speed of sound. The values of c' and v are given accurately by the theory of solids; they have been measured for many substances and may readily be measured for others. It is much more difficult to find $\bar{\lambda}$. We may estimate $\bar{\lambda}$ by taking the mean thermal conductivity of non-metallic solids as 0·01 cal/cm sec °C, v as 2×10^5 cm/sec, and c' as 0·5 cal/cm^3. Then $\bar{\lambda}$ is about 10^{-7} cm.

The thermal conductivities of glasses confirm this result. It is reasonable to assume that $\bar{\lambda}$ equals the mean distance between molecules for an amorphous body, since each molecule acts as a sharp discontinuity which produces large-angle scattering; scattering at thermal density fluctuations plays only a minor part. The concentration of such discontinuities is almost independent of temperature up to temperatures so high that the structure begins to become ordered. The speed of sound also varies very little with temperature. Therefore κ_1 for an amorphous solid should be roughly proportional to the specific heat, as is actually the case for glasses, especially at low temperatures for which c increases rapidly with temperature, while the scattering at thermal fluctuations is still very slight. However, although κ_1 rises with temperature on account of the rise in c, κ_1/c falls somewhat, because the thermal scattering becomes more important.

Measurements on neutron-irradiated crystalline quartz confirm the idea that phonons are scattered at discontinuities of atomic dimensions. The thermal conductivity (i.e. the mean free path) falls as the proportion of atoms displaced increases; the value for amorphous silica is reached when the integral flux is 10^{19}/cm^2. This corresponds to a 10-fold reduction in thermal conductivity.

We may find $\bar{\lambda}$ and the concentration of discontinuities by measuring κ_1, c' and v. The values of $\bar{\lambda}$ for glasses, amorphous silica, and heavily irradiated quartz are about $(7\text{–}10) \times 10^{-8}$ cm, which values are close to the mean distance between molecules of the glasses. Therefore equation (181) gives values of $\bar{\lambda}$ which correspond to the distances between

molecules. Crystalline bodies differ from amorphous ones only in the nature of the scattering centres, and hence in their values of $\bar{\lambda}$; the mechanisms are the same. In crystals $\bar{\lambda}$ is 10–100 lattice spacings, which is of the order of the mean free path for electrons.

It is now accepted that the basic physical concepts of Debye's theory are correct, but the exact form given to them must be changed to accord with quantum mechanical ideas. Elastic waves exchange energy only as phonons of energy $h\nu$ and momentum $h\nu/v$, where v is the speed of sound. The energy may change by $h\nu$, and the momentum by about $h/2\pi a$ (where a is the lattice parameter) when a phonon is deflected or absorbed.

Peierls has given the theory of thermal conductivity in quantum form; it appears that phonons are scattered only when not less than three collide. The thermal resistance is caused by a special form of phonon collision, in which the total quasi-momentum of the phonons is not conserved; the lattice absorbs or donates an amount of momentum $h/2\pi a$. The direction of the quasi-momentum is thereby reversed, or nearly so. Peierls calls these umklapp processes; in them the energy is conserved:

$$h\nu_1 \rightleftharpoons h\nu_2 + h\nu_3.$$

At high temperatures transverse waves play the main part in exchanging energy with the lattice; longitudinal waves take over this role at low temperatures. Inelastic collisions, in which the energy is increased or decreased by interaction with the lattice, play a secondary part in Peierls' theory.[57]

Scattering at impurities and lattice defects also affects the thermal conductivity. These processes become important at low temperatures, at which scattering by thermal fluctuations is slight, and at which umklapp scattering becomes comparable with scattering at the defects and impurities. These various processes contribute additively to the thermal resistance $1/\kappa_1$. Peierls gives the low-temperature thermal resistivity as

$$\frac{1}{\kappa_1} = AT^m \exp\left(-\frac{\theta}{2T}\right) + BNT^n - \frac{C}{LT^3} \qquad (182)$$

in which the first term represents the umklapp processes. The umklapp conditions imply that the sum of the wave vectors must change by $2\pi/a$; this serves to define the energies of the colliding phonons. The wave vector of any one phonon cannot exceed π/a; if the sum is to change by $2\pi/a$, one or two of the phonons must have momentum close to $h\pi/2a$. The number of such phonons is large above the Debye temperature ($T > \theta$), and so the condition $k \fallingdotseq \pi/2a$ does not restrict the number of scattering events. The situation is otherwise if $T < \theta$.

Now θ is related to the maximum frequency ν_0 of the thermal oscillations, i.e. to the angular frequency $\omega_0 = 2\pi\nu_0$:

$$\theta = \frac{h\nu_0}{k} = \frac{\hbar\omega_0}{k}.$$

However, ω_0 is approximately equal to π/a, where a is the lattice constant. The number of phonons having a frequency ω is given by Bose-Einstein statistics as proportional to

$$\frac{1}{\exp\left(\frac{\hbar\omega}{kT}\right)-1}$$

which reduces to $\exp(-\hbar\omega/kT)$ when $\exp(\hbar\omega/kT) \gg 1$. This gives $\exp(-\theta/2T)$ if ω equals $\omega_0/2$. The thermal resistance is proportional to the number of phonons that can suffer umklapp collisions, i.e. to $\exp(-\theta/T)$, and to the probability of an encounter between three phonons, which increases with temperature somewhat more rapidly than as the first power of T. Therefore the low-temperature thermal resistance caused by umklapp scattering can be put as

$$\left(\frac{1}{\kappa_1}\right)_{\text{umklapp}} = AT^m \exp\left(-\frac{\theta}{2T}\right) \quad \text{if} \quad T < \theta, \tag{182a}$$

where $2 > m > 1$. Now if $T \gg \theta, \hbar\omega/kT \ll 1$, so

$$[\exp(\hbar\omega/kT)-1]^{-1} \fallingdotseq 2kT/h\omega_0 \fallingdotseq 2T/\theta;$$

here we may take the probability of a suitable three-phonon encounter as temperature-independent. Therefore to a good approximation we have

$$\left(\frac{1}{\kappa_1}\right)_{\text{umklapp}} \propto \frac{T}{\theta} \quad \text{if} \quad T > \theta. \tag{183}$$

Pomeranchuk[58] has allowed for four-phonon encounters also, and concludes that the resistance must rise more rapidly with temperature (say as $T^{3/2}$ or $T^{5/4}$) when $T > \theta$. The measurements so far made favour a linear law, but not a directly proportional relation, i.e. they imply a law of the form

$$\left(\frac{1}{\kappa_1}\right)_{\text{umklapp}} = A\left(\frac{T}{\theta}-\frac{1}{3}\right). \tag{183a}$$

The second term in equation (182) arises from the impurities and defects. This term is proportional to the number N of such defects or other scattering centres encountered per unit path length. The smaller the ratio of wavelength to size of defect the greater the scattering. Now at temperatures well below θ the predominant phonons have wavelengths that exceed the dimensions of the atoms or lattice defects, so

such phonons are scattered only slightly. Phonons of shorter wavelength become dominant as the temperature rises, and these are scattered more strongly. Therefore the resistance caused by defects and impurities is proportional to N and increases slowly with temperature, as the second term in Peierls' formula shows:

$$\left(\frac{1}{\kappa_1}\right)_{\text{defects}} = BNT^p \tag{182b}$$

where $p < 1$. This expression is correct for temperatures near the Debye temperature. At low temperatures we have to allow for the increase in specific heat with temperature (c is proportional to T^3), and so we get

$$\left(\frac{1}{\kappa_1}\right)_{\text{defects}} = BNT^{(p-3)}.$$

Devyatkova and Stil'bans have measured the thermal resistance of sylvine crystals coloured by treatment with potassium vapour[59]; they measured the colour centre concentrations in terms of the absorption coefficients. They found that $1/\kappa_1$ is proportional to N at temperatures near to or above the Debye temperature.

Finally, the last term in equation (182) gives the resistance caused by reflection at the boundaries of grains or crystals in crystalline aggregates. This term is small at high temperatures, but at low temperatures, for which the scattering within the grains is negligible, we may take $\bar{\lambda}$ as being proportional to the linear size L of the grains. The specific heat is proportional to T^3 at low temperatures, while v varies little with temperature, so

$$\left(\frac{1}{\kappa_1}\right)_{\text{boundaries}} = \frac{C}{LT^3}. \tag{182c}$$

The same argument implies that the resistance caused by impurities has a factor T^q which is close to T^{-3} at low temperatures.

There is no simple method of deriving A, B, and C, so it is not possible to calculate the resistance of any particular material from Peierls' theory; the data show that equation (182) correctly predicts the shapes of the curves for single crystals (Fig. 117) and for powders at low temperatures.

Let us now try to estimate the resistance caused by phonon scattering for temperatures above θ. The mechanisms corresponding to equations (182b) and (182c) are here of secondary importance. The main causes of scattering in single crystals and in crystalline aggregates are the density fluctuations, which exceed the limits within which Hooke's law applies, and which thus make the oscillations nonlinear (anharmonic). Now c' and v can be taken as being independent of T when $T > \theta$. We shall make the following simplifying assumptions about $\bar{\lambda}$. Phonons are scattered

the more strongly the more vigorous the thermal motion and the more anharmonic that motion. We shall assume that the scattering probability (in the form $1/\bar{\lambda}$) is proportional to the energy u of thermal motion at $T > \theta$, at which temperatures the law of equipartition of energy is applicable. This energy U which has been accumulated by heating the body from absolute zero to T, is given by

$$U = \int_0^T c_v' \, dT;$$

the subscript v may be ignored, because solids have very small expansion coefficients. The anharmonicity is allowed for by a factor A', which we assume to vary only slightly with T; then κ (or, better, $1/\kappa$) is defined by the relation of U to T.

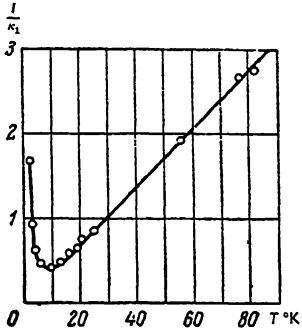


Fig. 117. Relation of lattice thermal resistivity to temperature for KCl.

The factor A' varies from one body to another, but we may expect that bodies whose binding forces and crystal lattices are of the same type will have values of A' that are similar or that vary regularly with the lattice constant and atomic weight. Now $1/\bar{\lambda}$ must increase with U, but this does not imply that the two are directly proportional. We shall, however, assume that this is actually so, since this assumption is correct for such cases as have been examined; then

$$\frac{1}{\bar{\lambda}} = A' \int_0^T c_v' \, dT. \tag{183b}$$

Our assumption is justified by the analogy with the scattering of electrons at thermal fluctuations in metals. The waves mainly responsible for heat transfer at $T > \theta$ are those whose wavelengths are of the order of atomic separations. Electrons near the Fermi level in metals have about that wavelength. It is thus reasonable to expect that phonon scattering will depend on the thermal motion in the way that electron scattering in

metals does. The two types of scattering, although different in nature, involve wavelengths of the same order.

The data for pure metals indicate that the resistance, and hence the reciprocal of the mobility, or the reciprocal of the mean free path, $(1/l)$, is proportional to the thermal energy at all temperatures:

$$\frac{1}{l} = D\int_0^T c_v' \, dT.$$

The factor D (for electrons) has a meaning distinct from that of A' (for heat waves); the first is determined only by the fluctuations in the electric fields, whereas the second is determined by the anharmonicity of the thermal motion. However, in both cases the dependence on temperature is caused by the thermal motion.

We assume that the mean free path of a phonon is given by an analogous formula; then equation (181) shows us that

$$(\kappa_1)_l = \tfrac{1}{3}c' v l A' \int_0^T c_v' \, dT.$$

The analogy with the electrical resistance becomes less suitable at low temperatures, since the effective wavelength of the thermal waves increases as T falls below θ, whereas the electron wavelength, which is defined by the Fermi level, scarcely varies with temperature down to absolute zero.

We may calculate $\int_0^T c_v' \, dT$ from Debye's specific heat formula, which often expresses the relation of c to T very well. The relation of $\int_0^T c_v' \, dT$ to T for $T > \theta$ is a straight line, which intersects the temperature axis at T equal to $\theta/3$. Therefore in this range of temperatures $\int_0^T c_v' \, dT$ is given by

$$\int_0^T c_v' \, dT = c'\left(T - \frac{\theta}{3}\right).$$

For $T > \theta$ we may also use the approximation $c = 3R$. Then

$$\int_0^T c_v' \, dT = 3R\left(T - \frac{\theta}{3}\right) \tag{184}$$

so that part of $(1/\kappa_1)$ which depends on thermal fluctuations is given by equation (181) as

$$\left(\frac{1}{\kappa_1}\right)_l = 3A' \frac{T - \theta/3}{v} \tag{185}$$

above the Debye temperature, as we indicated in equation (183a). Now in equation (185)

$$v = \sqrt{\frac{E}{\rho}} \tag{186}$$

where E is the elastic modulus and ρ is the density.

Several approximate expressions for θ have been proposed. Einstein's formula for the frequency limit of the oscillations gives

$$\theta = \frac{1{\cdot}8 \times 10^{-8}}{M^{\frac{1}{3}} \rho^{\frac{1}{3}} x^{\frac{1}{2}}} \tag{187}$$

where M is the molecular or atomic weight, ρ is the density and x is the compressibility (the reciprocal of the bulk modulus of elasticity).

Now equation (185) shows that $(1/\kappa_1)_t$ increases with T; the resistance for $T > \theta$ will be the greater the lower θ and the smaller v. Further, equations (186) and (187) indicate that $(1/\kappa_1)$ will increase with x (or with $1/E$), with ρ, and with M. Also A' must increase with the amplitude of the thermal motion for a given T, i.e. with x. Thus we expect that the heavier the atoms or groups in the crystal, and the weaker the binding forces, the larger will be the anharmonicity, and hence the lower will be the thermal conductivity.

The data on the relation of thermal conductivity to mean atomic weight confirm this. The non-metallic crystals and their compounds have thermal conductivities of about 0·1 cal/cm sec °C if we consider the lighter elements (B, C, Si, SiC); the elements in the middle of the table give values of about 0·01 in the same units, whereas the heaviest elements give values of only 0·001 or so. Hence κ_1 decreases as the atomic weight increases. The low thermal conductivities of sulphur, selenium and tellurium (which elements are of low atomic weight and appear to violate the rule) would appear to be caused by peculiarities in the structures, which contain rings or chains of atoms forming large molecules. These rings or chains presumably vibrate as a whole, which would make the effective molecular weight of say sulphur 32 times 8, i.e. 256, which would agree well with the low thermal conductivity (5×10^{-3} cal/cm sec °C).

There are few data suitable for checking the formulae thoroughly. Our measurements confirm that the thermal conductivity falls as M increases for series of analogous lattice types. Figure 118 gives our results and others' published data.[60] Here κ ranges from 0·45 to 0·002 cal/cm sec °C, so the values are put in logarithmic form. The atomic weights are exact for elements and are mean values $\overline{M}$ for compounds. The literature data relate to room temperature; ours relate to 280° K.

Figure 118 shows how the anharmonicity depends on the type of bonding. The top line 1 gives log κ for diamond-type lattices, which

have covalent bonds. Line 2 relates to compounds formed between elements of groups III and V, e.g. GaSb and InSb; the lattices resemble that of diamond, but the bonds are partly ionic. Line 3 is for compounds of oxides of group II elements; here the bonds are largely ionic. The two sub-groups differ in degree of ionic character, but the measurements do not reveal this difference. Compounds of groups I and VII (lines 4) give typical ionic crystals, though covalent bonds are still present. Here we find a difference between crystals whose ions have similar atomic weights (e.g. KCl) and crystals whose ions differ greatly. The thermal vibration spectra of these two types differ; in the first type the optical branch is only weakly developed, so the thermal conductivities are higher. All these features are reflected in Fig. 118. The fall in conductivity with increase in atomic weight is very evident.

Fig. 119a illustrates the effects of impurities on phonon scattering in terms of our measurements of $1/\kappa$ for Ge as a function of Si content, and for Si as a function of Sn content.[61] The curves give us the extra scattering caused by one impurity atom, i.e. the scattering cross-section. Interstitial impurities have larger scattering cross-sections than do substitutional impurities. Figure 119a applies to fairly small impurity concentrations; compounds whose structures are the same and whose lattice constants are similar often form a complete series of solid solutions. Such compounds are called isomorphous. We have made measurements on some series of such solid solutions. Figure 119b shows a typical result, for PbSe in PbTe, with $1/\kappa_1$ as a function of PbSe content (the contribution from the electrons has been calculated from the Wiedemann-Franz law and has been subtracted). In all cases we have found that the minimum thermal conductivity is close to that to be expected for an amorphous substance. In these cases $\bar{\lambda}$ ranges from one to two times the lattice constant. There is no reason to assume that the solid solution has at this point ceased to be crystalline, or that long-range order is absent. The X-ray diffraction patterns for the solid solutions are just as sharp as those for homogeneous crystals. The electron mobilities have high values, which correspond to undistorted lattices.

This difference between phonon and electron scattering may be ascribed to differences in wavelength; phonons have wavelengths of the same order as interatomic distances, whereas electrons have wavelengths some 20 times larger. Therefore phonons are scattered by inhomogeneities of atomic dimensions. Say, for example, that PbTe contains PbSe groups randomly distributed; then the latter act as scattering centres. (We would expect ordered solid solutions to have higher thermal conductivities than disordered ones, but no data are available.) The electron waves are long enough to be unaffected by defects of atomic size; the defects have to be of quarter wave-length size (i.e. 4–5 times

the atomic spacing). Solid solutions appear to be regularly arranged in this respect.

We may express the difference between phonon and electron scattering in the following way. Phonons are scattered at defects in the short-range order, whereas electrons are scattered at defects extending over several atomic diameters.

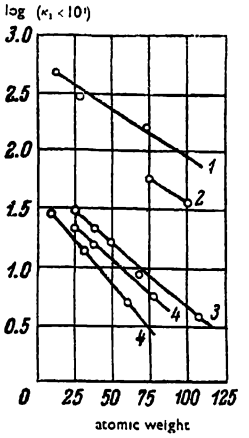


Fig. 118. Relation of thermal conductivity to mean atomic weight.

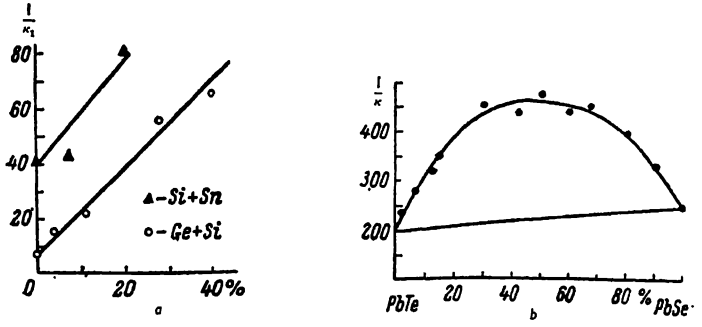


Fig. 119. (a) Relation of thermal resistivity to impurity content for Ge and Si. (b) Relation of thermal resistivity to PbSe concentration for solid solutions of PbSe in PbTe.

In obtaining the above results we derived the electron contribution to the thermal conductivity from the Wiedemann-Franz law and from equation (180). Our use of this law might appear to be justified by the results shown in Fig. 116. However, this conclusion is not to be extended; our results, and those shown in Fig. 116, relate to impurity conduction and to carriers of one sign. The electron contribution to κ has been

calculated from equation (180) and has been subtracted, to give

$$\kappa_1 = \kappa - A\left(\frac{k}{e}\right)^2 T,$$

which may be taken as the phonon conductivity.

Intrinsic conduction takes over from impurity conduction at high temperatures; we then get electrons in the conduction band and holes in the valence band in equal numbers. The heat conduction mechanism is then different. In an impurity conductor the electrons are prevented from diffusing to the cold end from the hot end by the electric field that is set up; therefore a temperature gradient causes no continuous flow of charge in an isolated rod if the carriers are all of one sign. The energy is transferred solely because the charges coming from the hot end have more energy than those returning from the cold end. However, if charges of both signs are present, a continuous flow can occur.

In 1940 Davydov and Shmushkevich showed that a flow of pairs of free carriers from regions where their concentration is high to regions where it is lower is accompanied by a gradual loss of those pairs by recombination. Any one such recombination releases energy ΔE_0, which becomes thermal energy. They found that the resulting thermal conductivity could be represented by a formula, which we may put in the form

$$\Delta\kappa = L(\sigma_+ + \sigma_-) + L\frac{2\sigma_+\sigma_-}{\sigma_+ + \sigma_-}\left(\frac{\Delta E_0}{2kT} + 2 + r\right)^2 \qquad (189)$$

where $L = A(k/e)^2 T$ is the coefficient of σ in equation (180), in which A equals $(2+r)$; σ_+ and σ_- are the electrical conductivities assigned respectively to the holes and electrons, and r is the exponent in the formula

$$l \propto \epsilon^r.$$

r is zero for atomic lattices with covalent bonds. The first term in equation (189) gives us the Wiedemann-Franz law, whereas the second gives us the extra thermal conductivity caused by the diffusion of electron-hole pairs.[10]

In 1952 Devyatkova found that κ_2 for lead telluride was higher than the value predicted by equation (180) at temperatures such that intrinsic conduction was apparent.[55] Gul'tyaev extended these measurements to still higher temperatures, at which intrinsic conductivity was dominant; he found even larger deviations from equation (180). The additional conductivity was calculated from

$$\Delta\kappa = \kappa - \kappa_1 - L\sigma.$$

The phonon contribution κ_1 was deduced by extrapolating the straight line (Fig. 120) to high temperatures. This extrapolation is justified,

because an insulator shows a linear relation of thermal resistance to temperature up to its melting point.

Figure 121 shows Devyatkova's and Gul'tyaev's results for PbTe. The intrinsic conductivity becomes appreciable at about 300° K, and κ_1 predicted by equation (180) begins to exceed the value found by

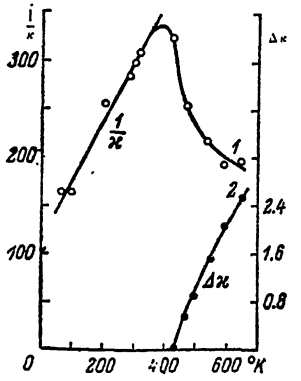


Fig. 120. (1) Relation of thermal resistivity to temperature for Bi_2Te_3, (2) relation of increase in thermal conductivity to temperature for Bi_2Te_3.

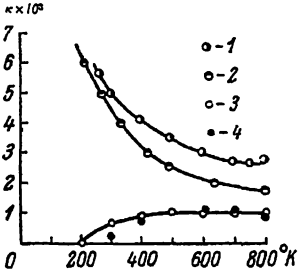


Fig. 121. Comparison of thermal conduction mechanisms.
(1) Experimental curve,
(2) $K_1+L\sigma$,
(3) $\Delta K = K-K_1-L\sigma$,
(4) ΔK exciton.

extrapolating the straight line for $1/\kappa$. The effect is still more marked with Bi_2Te_3, as Fig. 120 shows. In 1954 Busch found the same effect for InSb.[62] Thus we may assume that the thermal conductivity is increased if carriers of both signs are present.

However, the extra conductivity predicted by

$$\Delta\kappa = \kappa - \kappa_1 - L\sigma$$

does not always agree with equation (189), which is due to Davydov and Shmushkevich. In some cases, e.g. in Busch's work on InSb, for which

the electrons differ greatly in mobility from the holes, the factor $\sigma_+\sigma_-/(\sigma_+ + \sigma_-)$ is very small. The measured $\Delta\kappa$ is many times larger than that predicted by equation (189). The agreement is better with PbTe and Bi_2Te_3. However we must note that the process Davydov and Shmushkevich considered is not the only one possible in the region in which intrinsic conduction is appreciable. There are several other processes that can raise the conductivity above the value predicted by equation (189) if the conduction is of mixed type. We have mentioned excitons already; these are excited lattice states in which the electrons remain bound to the holes but the pairs can diffuse through the lattice. Excitons are produced when photons are absorbed; the energy needed is less than ΔE_0 (the width of the forbidden band), but the electron cannot separate from the hole, so the pair can carry no current. However, excitons can diffuse down a concentration gradient.

We may expect that thermal motion can also produce excitons, since the energy $\Delta E'$ needed is less than ΔE_0. The ratio between the number N of electrons in exciton states and the number of unexcited electrons in the valence band is $\exp(-\Delta E'/kT)$; the effective electron concentration in the valence band may be taken as $2(2\pi m^* kT)^{\frac{3}{2}}/h^3$. Hence N is given by

$$N = \frac{2(2\pi m^* kT)^{\frac{3}{2}}}{h^3} \exp\left(-\frac{\Delta E'}{kT}\right).$$

The excitons are destroyed when they have diffused from a region in which their concentration is high to one in which it is low, and each releases energy $\Delta E'$. Pikus[40] has shown that in this case the extra thermal conductivity is given by

$$\Delta\kappa_{exc} = ND\left(\frac{\Delta E'}{kT} + 2\right)^2 \tag{190}$$

where D is the diffusion coefficient for excitons. Ansel'm states that this coefficient is similar to the diffusion coefficients for holes and electrons, or much larger if the carriers are equal in mobility.[63]

The conductivity may also be increased by momentum transferred from the electron-hole and exciton fluxes to the phonons. The resulting phonons carry energy from the hot to the cold end.

Another form of thermal energy transport which is possible for semiconductors is at present under discussion; this form is one in which the free electron assembly and the corresponding positive ions have a thermal motion which is the stronger the higher the temperature. Holes must show a similar relation to negative ions. These oscillations give rise to waves that propagate along the direction of the temperature

gradient; they may carry energy and so may increase the thermal conductivity.

We should not neglect electromagnetic radiation since the density of this increases as T^4. This radiation carries heat down the temperature gradient since it is absorbed as it travels. The extra conductivity caused by this radiation has been given by Henzel as

$$\Delta\kappa_{\text{radiation}} = \frac{16}{3}\frac{a\epsilon}{k}T^3 \tag{191}$$

where ϵ is the optical equivalent of the dielectric constant (the square of the refractive index for short wavelengths), k is the absorption coefficient for those wavelengths, which must depend on the concentrations of the free electrons and holes, and a is the factor in the relation $U = aT^4$ for the radiation energy density of a black body.

Finally, the free electrons, if present at high concentration, may reduce the thermal conductivity because they act as scattering centres for the phonons. It is found that this effect becomes appreciable at concentrations of $10^{20}\,\text{cm}^{-3}$.

Of all these processes the only one for which exact calculations are possible is that of the free electron-hole pairs, which is covered by equation (189). We can estimate the relation of $\Delta\kappa_{\text{exc}}$ to T from equation (190). The effects of radiation, which are expressed by equation (191), become appreciable only above 700° K.

Let us now consider the thermal conductivities of crystalline aggregates. The calculations above make no allowance for scattering at grain boundaries; they are correct for crystals in which the speed of sound is the same in all directions. The boundaries between crystals then cannot cause appreciable reflection of waves whose wavelengths exceed the width of the gaps between crystals. This is not true for anisotropic crystals, in which the speed of sound and the thermal conductivity may vary greatly with the crystallographic direction. The surface representing the thermal conductivity as a function of direction is an ellipsoid (see Fig. 60). The differing acoustic impedances on the two sides of a junction between crystals whose orientations differ must cause the waves to be reflected. Therefore we expect that a fine-grained untextured aggregate of such crystals (e.g. of quartz or graphite) will show a relation of low-temperature thermal conductivity to crystal size L more marked than that given by equation (182b) for an isotropic substance. Reflections will also occur at the boundaries between grains in a pressed powder if the contact is poor. This extra thermal resistance has in fact been detected in crystalline materials at low temperatures; it depends on the grain size of the pressed powder. Casimir's calculations on the thermal resistance of a powder agree well with experiment.

Thermoelectricity:[64] discovery of thermoelectric phenomena

In 1822–1823 Seebeck described in the Reports of the Prussian Academy of Sciences a phenomenon which he named "the magnetic polarisation of metals and ores produced by a temperature difference". It is obvious from the description of his experiments that Seebeck had discovered thermoelectric currents arising in a closed circuit made of different conductors at different junction temperatures.

This was an epoch when the discovery by Oersted of the effect of an electrical current on a magnetic needle was followed by a series of investigations by Ampère, Biot, Savart, Laplace, and others who elucidated the interaction between electric currents and magnetic fields; Ampère's molecular current hypothesis reduced the source of magnetic phenomena to electrical currents. The science of electricity and magnetism grew from Monday to Monday (the day of the week on which the Paris Academy of Sciences used to sit); physicists, mathematicians and chemists played an active part in its development.

This scientific development spread from Paris to Switzerland and Germany, and then to England, where it was enriched by Faraday's discovery of electromagnetic induction.

When Seebeck's experiments are considered without prejudice there can be no doubt that the phenomena observed by him were caused by the electric current (the disappearance of the magnetic effects whenever the current was cut off by the introduction of a non-conducting layer, the quantitative relationship between the magnetic field produced by the current and the observed deflections of the magnetic needle, etc.). However, Seebeck not only rejected this natural explanation of his results but fought actively against it for several years after, accusing the exponents of the thermoelectric current theory of following a fashion started by Oersted's discovery.

An explanation of this dogmatic attitude of Seebeck towards his discovery may be found in the concluding part of his large treatise, in which he makes an attempt to relate the Earth's magnetism to the temperature difference between the equator or a range of southern volcanoes and the polar ice cap. It appears that this hypothesis was closer to his heart than the discovery of a further source of electric current. The value of Seebeck's paper does not, of course, lie in this hypothesis (he himself showed that a temperature difference without a closed electrical circuit does not produce any magnetic effects), nor in the attempt to deny the existence of an electric current, but in the copious experimental material accumulated by him and covering a great variety of solid and liquid metals, their alloys, minerals, and semiconductors. Seebeck's mistake had a positive effect: in order to disprove the electrical origin of thermoelectric currents he compared, for a great

variety of materials, the contact potential (Volta) effect with the effect of a temperature difference on a magnetic needle, and demonstrated the difference between these phenomena.

The extensive thermoelectric series compiled by Seebeck is, even now, of interest. Using modern notation (α = thermoelectric power, and σ = electrical conductivity), the Seebeck series is in the order of magnitude of the product $\alpha\sigma$ and not the quantity $\alpha^2\sigma/\kappa$ (where κ is the thermal conductivity) which, as will be seen later on, characterises the thermoelectric properties of materials. The thoroughness of Seebeck's investigations may be illustrated, for example, by the fact that in an article by Telkes published in an American journal 125 years later the best couple was given by ZnSb and PbS, which were the first and the last members in Seebeck's series.

Seebeck, 1822	Justi, 1948		Meisner, 1934/5			
			metals		semiconductors	
PbS	Bi	−80	Bi	−70	MoS	−770
Bi	Constantan	−39	Ni	−18	ZnO	−714
Bi amalgam	Co	−21	Co	−17·5	CuO	−696
Ni	Ni	−20	K	−12	Fe_2O_3 (400°C)	−613
Co	K	−14	Pd	−6	FeO	−500
Pd	Pd	− 8	Na	−4·4	Fe_3O_4	−430
Pt No. 1	Na	− 7	Hg	−3·4	FeS_2	−200
U	Pt	− 5	Pt	−3·3	MgO_3H_2	−200
Au No. 1	Hg	− 5	Al	−0·6	SnO	−139
Cu No. 1	C	−3·5	Mg	−0·4	Fe_2O_3 (50°C)	−60
Rh	Al	−1·5	Pb	−0·1	CdO	−41
Au No. 2	Mg	−1·5	Sn	+0·1	CuS	−7
Ag	Pb	−1·0	Cs	+0·2	FeS	+26
Zn	Sn	−1·0	W	+1·5	CdO	+30
C	Cs	−0·5	Ir	+2·2	$FeTiO_3$	+140
Cu No. 3	Manganin	+0·5	Ag	+2·4	NiO	+240
Pt No. 4	Ir	+1·0	Rh	+2·5	Mn_2O_3	+385
Cd	Rh	+1·0	Cu	+2·6	Cu_2O	+474
Steel	Zn	+1·5	Au	+2·7	Cu_2O	+1000
Fe	Ag	+1·5	Cd	+2·8	CuO	+1120
As	Au	+1·5	Zn	+2·9	Cu_2O	+1150
Sb	Cu	+2·0	Mo	+5·9		
ZnSb	W	+2·5	Fe	+16		
Te	Cd	+3·5	Sb	+35		
	Mo	+6·5	Te	+400		
	Fe	+12·5	Se	+1000		
	Sb	+42				
	Si	+44				
	Te	+49				

A comparison of the qualitative Seebeck series with modern measurements of the thermoelectric power of metals shows good agreement. The divergence between Seebeck's series and those of Justi and Meisner is not larger than the divergence between the latter two series. We are

reproducing below the aforementioned thermoelectric series (the thermoelectric power in Justi's and Meisner's series is expressed in μV/deg).

Had Seebeck followed his discovery with an attempt to use a thermocouple to generate electrical energy, he could have obtained, using the first and last terms in his series, an efficiency of the order of 3%, i.e., as much as that of the best steam engines at that time.

It must also be noted that Seebeck did not fail to pay attention to phenomena caused by a temperature difference in homogeneous materials and that he estimated qualitatively thermoelectric relationships which were rediscovered 30 years later by Thomson.

Twelve years after Seebeck's discovery, Peltier, a watchmaker, published an article on temperature anomalies observed in the vicinity of the boundary between two different conductors when a current was passing through them. The phenomenon first observed by Peltier, and therefore given the name of Peltier effect, consists in reality of the generation or absorption (depending on the direction of the current) of heat, at a rate Q, at the junction between two different conductors when a current J flows through them:

$$Q = \Pi J$$

where Π is the Peltier coefficient.

The Peltier effect is very closely related to the Seebeck effect, and the coefficient Π to the thermoelectric power α:

$$\Pi = \alpha T$$

where T is the absolute temperature of the junction, so that $Q = \alpha TJ$. A temperature difference produces an electric current in a closed circuit made of different materials, whilst, in turn, a current flowing through such a circuit produces a temperature difference. It is interesting that Peltier did not notice this connection in spite of the fact that he carried out his experiments using a thermoelectric circuit. In these experiments, the Seebeck effect was used solely for secondary purposes, namely as a source of weak currents.

The anomalies discovered by Peltier were found to be the more pronounced the larger the thermoelectric powers of the metals involved; they were particularly conspicuous at the junction of bismuth with antimony. Peltier was, however, looking for, and finding, in his experiments the confirmation of a preconceived idea that the universal law of heat generation by current, namely the Joule-Lenz law, is only valid for strong currents. Peltier thought that in the case of the weak currents produced by a thermocouple the individual properties of the metals began to have an effect. He did not find a confirmation of this hypothesis in the bulk properties of the conductors, but the anomalies at the junctions

were affected by the nature of the metal (in reality by its thermoelectric properties). Peltier did not notice this and sought an explanation in the hardness or softness of the metal or its electrical conductivity, and when the facts did not come up to his expectations, for example in the case of bismuth, the thermal anomalies of which were attributed by Peltier to a low electrical conductivity, he refused to believe the measurements.

A few more years passed, during which Becquerel and other investigators tried to explain the true nature of the Peltier phenomenon, until in 1838 the St. Petersburg Academician Lenz put an end to all doubts with a simple experiment. Lenz placed a droplet of water in a dent at the junction of bismuth and antimony rods. The droplet turned into ice when the current was reversed. It is known that to freeze 1 gm of water it is necessary to remove 80 cal, whilst in melting 1 gm of ice the same amount must be supplied. It thus became obvious that, depending on its direction, an electric current generates or absorbs heat at the junction of two conductors.

Thermoelectric phenomena did not attract attention amongst physicists. They were obscured by the striking phenomena of electromagnetism which led Faraday to the discovery of electromagnetic induction. Physics was moving towards the generalising laws of Maxwell's theory, and engineering towards electrical machinery. Thirty years had passed since Seebeck's discovery when, owing to the evolution of thermodynamics, all types of energy conversion became of interest, including the conversion of thermal and electrical energy in the Seebeck and Peltier effects.

This, in fact, was the approach adopted by Thomson, one of the founders of thermodynamics. Thermodynamic analysis of the Seebeck and Peltier effects led him not only to the establishment of the relationship between these two processes which has already been mentioned above, but also to the discovery of a third phenomenon, the Thomson effect, consisting of the generation or absorption of heat q by the passage of a current through a homogeneous conductor in which there is a temperature gradient $\partial T/\partial x$:

$$q = \tau J \frac{\partial T}{\partial x};$$

τ became known as the Thomson coefficient; as will be seen later it is related to the coefficients α and Π.

It should be remembered that the Thomson effect was earlier experimentally observed and investigated by Seebeck.

In 1857 Thomson published his theory of the thermoelectric phenomena in anisotropic crystals; however, only in the 20th century was it shown by Bridgman that a change of the direction of the current

in a crystal is connected with the absorption or generation of heat, as in the case of the Peltier effect. Such an internal Peltier effect was called the Bridgman effect.

In 1920–21, almost one hundred years after Seebeck's discovery, Benedicks reported the discovery of two more thermoelectric effects depending on the value of the temperature gradient and not on the temperature difference between the ends of the conductors. When the ends of a conductor are at the same temperature, but if there is a warmer or cooler portion in between them, then, according to Benedicks, a potential difference will appear between the ends of the conductor when the warm (or cool) portion is not symmetrically placed, i.e. is closer to one of the ends. A careful check of the Benedicks effect by other investigators did not confirm its existence. It may be presumed that even if this effect exists it is negligible.

The phenomena enumerated above exhaust the field of thermoelectricity, if phenomena arising in a magnetic field, generally referred to as thermomagnetic phenomena, are not included.

Thermocouples have been used for temperature measurement for a long time, but the problem of their power application became of importance only with the growing demand for sources of electrical energy.

In 1885 Rayleigh considered the problem of the efficiency of a thermoelectric generator and calculated it, though somewhat inaccurately. The problem of a thermoelectric electrical energy generator was again taken up in 1909 by Altenkirch, this time, on the whole, correctly. In 1910 Altenkirch investigated the problem of thermoelectric cooling and heating. However, since the only conductors technically known at that time were metals, such devices were found to be uneconomical. It is true that a few thermoelectric generators (by Gülcher, Coblentz, and others) made their appearance, but they did not find wide application since their efficiency did not exceed 0·6% and sometimes not even 0·1%.

The position on the electrical engineering front (and, therefore, also the subjects of physical investigations) underwent a radical change with the advent of semiconductors. The prospects of attaining appreciable efficiencies in thermoelectric generators became more favourable. In conjunction with the problems of industrialisation of the Soviet Union, at the beginning of the first five-year plan, I outlined in 1929 the advantages of thermoelectric generators made of semiconductors and calculated that their efficiency could reach 2·5 to 4% with prospects of an appreciable increase in the future. In fact, in 1940 Maslakovets described a thermoelement, made of lead sulphide with an excess of lead and sulphur, with an efficiency of the order of 3%.

In 1947 Telkes claimed an efficiency of 7% for her thermoelements; however the thermoelements which she described in detail in 1954 had an efficiency not exceeding 3·3%.

In 1953 Justi published design calculations for refrigerators, utilising metallic and semiconducting materials and concluded, incorrectly, that there is a limit of 25° C to the temperature difference produced by the thermoelectric cooling. Later that year, Goldsmid and Douglas reported a temperature drop of 26° C.

It seems that all the energetic applications of thermoelectricity: (1) thermoelectric generators, (2) thermoelectric refrigerators, and (3) thermoelectric industrial and domestic heating devices, are not only of great theoretical interest, but present real prospects of practical application on the basis of the modern science of semiconductors. Moreover, some previously unforeseen applications have arisen, such as the construction of thermoelectric sound and ultrasonic generators, vacuum thermoelements, etc.

Three thermoelectric effects have been established experimentally:

1. The appearance of a thermoelectric voltage dV in an open circuit consisting of two different conductors when there is a temperature difference dT between their ends:

$$dV = \alpha_{1-2}\, dT \tag{192}$$

(α_{1-2} is known as the thermal emf coefficient, or thermoelectric power, between the given conductors; α may depend on the temperature);

2. The generation of heat dQ at the boundary between two conductors when a current, dJ, flows through them:

$$dQ = \Pi_{1-2}\, dJ \tag{193}$$

(change of direction of the current leads to a change from the generation of heat to the absorption of heat and vice versa);

3. Generation of heat dq by a current J in a portion of the conductor with a length dx in the presence of a temperature gradient $\partial T/\partial x$:

$$dq = \tau J \frac{\partial T}{\partial x}\, dx. \tag{194}$$

The relationship between α_{1-2}, Π_{1-2} and τ follows from the laws of thermodynamics.

We shall consider a closed circuit made up of two conductors with a temperature difference dT between the junctions. According to the law of conservation of energy we can equate the algebraic sum of all types of energy generated in the circuit in a unit time to zero under steady-state conditions.

The Peltier heat in watts generated at one junction will be denoted by $\Pi_{1-2}\,dJ$, whilst the heat absorbed at the other junction will be denoted by $\Pi_{2-1}\,dJ$; we have

$$\Pi_{2-1} = \Pi_{1-2} + \frac{\partial \Pi_{1-2}}{\partial T}\,dT.$$

The Thomson heat generated in one second in one conductor is $\tau_1\,dJ\,dT$, whilst that absorbed in the other conductor is $\tau_2\,dJ\,dT$. The electrical power consumed in the circuit is $\alpha\,dT\,dJ$.

In addition to these three processes, we have to take into account the irreversible processes of thermal conduction and generation of heat in accordance with the Lenz-Joule law. The first of these processes does not affect, under steady-state conditions, the total amount of energy, and leads only to a transfer of heat from one portion of the circuit to another. In the second process energy is generated at a rate proportional to the square of the current dJ, but this energy remains in the circuit in the form of heat.

Consequently, the equation of conservation of energy in the circuit can be written down in the form

$$\frac{\partial \Pi_{1-2}}{\partial T}\,dT\,dJ + (\tau_1 - \tau_2)\,dJ\,dT = \alpha_{1-2}\,dT\,dJ$$

or

$$\frac{\partial \Pi_{1-2}}{\partial T} + \tau_1 - \tau_2 = \alpha_{1-2}. \tag{195}$$

Another relationship between the coefficients α, Π, and τ may be derived from the equation of the second law of thermodynamics, which is, however, only applicable to reversible processes.

All three thermoelectric phenomena can, as a first approximation, be regarded as reversible; their sign changes both with a change of the sign of the temperature difference and with a change in the direction of the current.

However, the irreversible phenomena of heat conduction and Joule heat generation will also inevitably take place. Whilst the Joule heat, as we shall see later, may be neglected for extremely small currents, the heat transferred by thermal conduction is of the same order as the Peltier heat and sometimes much greater.

Thomson assumed that the thermoelectric phenomena are not basically related to the heat conduction process and the generation of heat by the current. The thermal conductivity of a given material and its resistivity, which govern the irreversible processes in thermocouples, do not determine the Seebeck, Thomson, and Peltier effects. These last three phenomena may be considered in isolation, assuming that the thermal conductivity and the resistivity are arbitrarily low and neglecting irreversible processes as incidental.

These considerations cannot be regarded as a convincing proof. More recently Onsager stated the conditions under which such a division of processes into reversible and irreversible is permissible. It appears that Onsager's conditions are satisfied both by metals and semiconductors. This cannot, however, be regarded even now as definitely established.

Following Thomson, we shall assume, however, that the reversible part of thermoelectric processes may be considered separately from the irreversible processes and we shall then check the resulting conclusions. Agreement between these conclusions and experimental results may be regarded as some confirmation of the underlying assumptions. Since we shall consider below only the theoretical conclusions, the theory will be regarded here merely as a way of deriving these results.

Thus, we shall apply the second law of thermodynamics to thermoelectric processes in a thermocouple with hot and cold junction temperatures T_1 and T_0 respectively. The hot junction receives an amount of heat $\Pi_1 J$, and the cold junction gives up an amount of heat $\Pi_0 J$; at the same time an amount of heat $\tau J\,dT$ is generated at a temperature T in each portion of the conductors over which there is a temperature difference dT.

Reversibility of the processes leads to the condition that the total change of entropy is equal to zero:

$$\frac{\Pi_1}{T_1}J-\frac{\Pi_0}{T}J+J\int_T^{T_1}\frac{\tau_2-\tau_1}{T}\,dT=0.$$

Dividing this equation by J and differentiating it with respect to T, we obtain

$$\frac{d\frac{\Pi}{T}}{dT}+\frac{\tau_2-\tau_1}{T}=0,\quad \frac{T\frac{d\Pi}{dT}-\Pi}{T^2}=\frac{\tau_1-\tau_2}{T},$$

whence

$$\frac{d\Pi}{dT}-\frac{\Pi}{T}=\tau_1-\tau_2. \tag{196}$$

Equation (196) together with equation (195) gives

$$\alpha=\frac{\Pi}{T}. \tag{197}$$

Each of the quantities in expression (197) may be separately measured, thus making it possible to check this expression, obtained by separating reversible from irreversible processes in the thermocouple. Experimental material so far accumulated has always confirmed equation (197) within the limits of experimental accuracy. It will therefore, be regarded as an

experimental fact. This equation gives

$$\frac{d\alpha}{dT} = \frac{T\frac{\partial \Pi}{\partial T} - \Pi}{T^2} = \frac{1}{T}\frac{\partial \Pi}{\partial T} - \frac{\Pi}{T^2} = \frac{1}{T}\left(\frac{\partial \Pi}{\partial T} - \frac{\Pi}{T}\right).$$

According to equation (196) the expression in brackets is equal to $(\tau_1 - \tau_2)$. We have therefore

$$\frac{d\alpha_{1-2}}{dT} = \frac{1}{T}(\tau_1 - \tau_2). \tag{198}$$

On the left hand side of equation (198) is a quantity relating to the boundary between the two conductors, and on the right hand side a difference between two terms, each of which depends on the properties of only one conductor.

A question can, therefore, be posed as to whether α_{1-2} may be regarded as a difference between two quantities, α_1 describing the first conductor and α_2 describing the second conductor. Integrating equation (198) we obtain

$$\alpha_{1-2} = \int_0^T \frac{\tau_2}{T} dT - \int_0^T \frac{\tau_1}{T} dT,$$

where the lower limit of the integral corresponds to the temperature of absolute zero, at which both α_{1-2} and $d\alpha_{1-2}/dT$ are equal to zero. It is natural to set

$$\alpha_{1-2} = \alpha_2 - \alpha_1 = \int_0^T \frac{\tau_2}{T} dT - \int_0^T \frac{\tau_1}{T} dT,$$

$$\alpha_2 = \int_0^T \frac{\tau_2}{T} dT, \quad \alpha_1 = \int_0^T \frac{\tau_1}{T} dT. \tag{199}$$

In order to determine α_1 and α_2 in practice it would be necessary to know $\tau(T)$ from the absolute zero to the specified temperature. Experimental measurement of the Thomson effect over such a wide temperature range is difficult and so far has only been accomplished for the normal silver alloy. The difference $(\alpha_2 - \alpha_1)$ and its temperature dependence have been determined frequently for a wide variety of materials.

It can be readily shown that by measuring, at a certain temperature T_0, α_{1-2}, α_{2-3}, $\alpha_{3-4}, \ldots, \alpha_{(n-1)-n}$, we obtain $\alpha_{1-2} + \alpha_{2-3} + \alpha_{3-4} + \ldots + \alpha_{(n-1)-n} = \alpha_{1-n}$. In fact, since all the junctions 1—2, 2—3, 3—4 ... are at the same temperature T_0, any change of the emf of the couple $1-n$ by connecting into the circuit intermediate conductors 2, 3, ... would require the conversion of heat from the surroundings into electrical energy without any compensation, which contradicts the second law of thermodynamics. The independence of the value of α_{1-n} from the effect of the

intermediate conductors forms the basis for setting up a thermoelectric series. Experiment has always confirmed this relationship.

If we could ascribe to some substance a definite value $\alpha = \alpha_0$ we would then be able to determine the absolute as well as the relative values of the thermal emf coefficient of all substances. Equation (199) makes this possible once $\int_0^T (\tau/T)\,dT$ has been measured for one substance only; the accuracy of such a measurement is, however, low, although numerous investigators (Borelius, Keesom, and others) have made attempts to determine $\tau(T)$ close to the absolute zero. Within experimental accuracy (about 1 μV/deg) α may be regarded as zero for lead and tin. Comparison with the Hall effect affords another method for determining the absolute values of α. Although the conditions for α and R to be zero in mixed semiconductors differ slightly, it may be assumed with sufficient accuracy that $\alpha = 0$ when $R = 0$. R is found to be approximately zero for Sn and Pb, so the results given by the two methods coincide.

Using equations (197) and (198) it is possible to express the Peltier coefficient Π and the Thomson coefficient τ in terms of α and $d\alpha/dT$. In particular, when $d\alpha/dT = 0, \tau_1 = \tau_2$ and therefore the quantities of Thomson heat generated in the two branches of the thermocouple are equal and opposite in sign. The Thomson heat generated in a closed circuit is

$$q = (\tau_1 - \tau_2) J (T_1 - T_2),$$

and the Peltier heat is $$Q = \alpha J T.$$

The higher $(T_1 - T_2)$ as compared with T, the greater is the relative part played by Thomson heat.

It would appear that purely thermodynamic considerations make possible not only the establishment of a relationship between Π, τ, and α, but also the determination of the actual value of α. In fact, let us consider a junction between two electronic conductors through which passes a certain number of electrons, dn, at slow rate, so that the current J is infinitesimally small; the temperature of the whole circuit is assumed to be constant and equal to T. In such a circuit there is no heat conduction or Joule heat loss and the whole circuit may be regarded as a system in equilibrium in which all processes change their sign with the change of the sign of J.

According to the second law of thermodynamics the amount of heat generated at the junction is

$$dQ = T dS = T(S_1 - S_2)\,dn,$$

where S_1 and S_2 are, respectively, the entropies per electron for the first and the second material. From the definition of the Peltier

coefficient we have $$dQ = \Pi_{1-2}\, e\, dn.$$

Hence $$S_1 - S_2 = \frac{\Pi_{1-2}}{T} = \alpha_{1-2}. \tag{200}$$

It might have been expected that the thermal emf α at the boundary between two electronic conductors could be determined as a difference of entropies of 1 coulomb of electricity in the two conductors. This assumption is, however, not entirely correct: in the equilibrium state the electron velocity distribution follows a certain statistical law. In a stream of electrons the velocity distribution changes; for example, fast electrons pass from one conductor into another in larger numbers than slow electrons. Electrons which remain in the first conductor restore the statistical equilibrium which brings about a further change of entropy. Therefore the entropy flux from one substance into another is not equal to the difference of the entropies in each of the substances before the exchange of the electrons. Equation (200) may only be regarded as an approximation. Calculation of the exact value of α_{1-2} is, therefore, only possible starting with the consideration of the mechanism of electron transfer.

Let j denote the current density of electrons through the boundary between two substances; this current consists of currents of electrons with all values of kinetic energy ϵ:

$$j \propto \int_0^\infty j(\epsilon)\, d\epsilon.$$

The average value of the energy of electrons in the current is

$$\bar{\epsilon} = \frac{\int_0^\infty \epsilon j(\epsilon)\, d\epsilon}{\int_0^\infty j(\epsilon)\, d\epsilon}.$$

Let $f_0(\epsilon)$ denote the equilibrium energy distribution function of the electrons and $l(\epsilon)$ the free path length of an electron with an energy ϵ; the current j can then be represented in the form

$$j = C \int_0^\infty \frac{df_0(\epsilon)}{d\epsilon} l(\epsilon) \epsilon\, d\epsilon.$$

This expression is equally true for a current of gaseous molecules. Substituting it into the equation for $\bar{\epsilon}$ we obtain

$$\bar{\epsilon} = \frac{\int_0^\infty \frac{df_0(\epsilon)}{d\epsilon} l(\epsilon)\, \epsilon^2\, d\epsilon}{\int_0^\infty \frac{df_0(\epsilon)}{d\epsilon} l(\epsilon)\, \epsilon\, d\epsilon}.$$

If S is the flow of entropy with the passage of one electron, then the flow of entropy S/e corresponding to 1 coulomb of electrical charge, on which the value of α depends, will be described by the expression

$$\alpha = \frac{S}{e} = \frac{1}{e}\left(\frac{\bar{\epsilon}}{T} - \frac{\mu}{T}\right) = \frac{k}{e}\left(\frac{\bar{\epsilon}}{kT} - \frac{\mu}{kT}\right).$$

Using the notation $\epsilon/kT = \epsilon^*$ and $\mu/kT = \mu^*$, we have

$$\alpha = \frac{k}{e}\left[\frac{\int_0^\infty \frac{df_0(\epsilon^*)}{d\epsilon^*} l(\epsilon^*)\,\epsilon^{*2}\,d\epsilon^*}{\int_0^\infty \frac{df_0(\epsilon^*)}{d\epsilon^*} l(\epsilon^*)\,\epsilon^*\,d\epsilon^*} - \mu^*\right]. \tag{201}$$

The Peltier heat is

$$\Pi = \alpha T = \frac{1}{e}(\bar{\epsilon} - \mu). \tag{202}$$

These are the most general expressions for the thermal emf and the Peltier coefficient which are valid both for degenerate and non-degenerate semiconductors and metals.

At the beginning of this century, the theory of metals was based on the concept of the presence in each metal of a certain concentration of free electrons n, different for differing metals, moving at random within the metal in a similar way to gas molecules. According to these theories (Rieke, Drude, Lorentz and Debye) electrons were governed by the same Boltzmann statistics as gaseous molecules. As in a gas their average translation energies were regarded as equal to $3/2.kT$ where k is the Boltzmann constant and T the absolute temperature.

The thermal emf V in an open circuit consisting of two metals, calculated according to this theory, is

$$V = \frac{k}{e}\int_{T_1}^{T_2} \log\frac{n_2}{n_1}\,dT. \tag{203}$$

It can be readily calculated that, e.g., when

$$n_2/n_1 = 2{\cdot}7,\ V = 86 \times 10^{-6}(T_2 - T_1) \text{ volt},$$

and $\alpha = 86\,\mu V/\text{deg}$, which is much higher than the experimentally observed value of α in cases where one would expect that $n_2/n_1 > 2{\cdot}7$. For the majority of metals α does not exceed a few μV/deg.

The quantum theory of metals developed in 1927–1928 by Frenkel' in one form and by Sommerfield in another form eliminated this discrepancy as well as many others.

According to the theory of Sommerfeld who applied Fermi-Dirac quantum statistics to electrons in metals, the thermal emf between two metals is to a first approximation (taking into account only linear terms)

equal to zero. Only the next approximation leads to a certain finite, although very small, value of α. All this is not surprising. Let us consider a metallic rod, the ends of which are at different temperatures. Since the temperature does not change the concentration of electrons n in the metal but brings about only a slight redistribution of their thermal agitation velocities, it is obvious that a large thermal emf cannot arise in such a metal.

The state of semiconductors in the quantum theory is similar to that of the classical metal considered at the beginning of the 20th century. Some of the electrons in semiconductors are found to be free. The concentration of free electrons in semiconductors varies but is usually so much smaller than in metals that the Boltzmann classical statistics are applicable. A rise of temperature leads to a change in both the concentration of free electrons and their kinetic energy which, as in the case of gas molecules, may be equated to $3/2.kT$.

Let us consider a rod of semiconductor in which there is a temperature gradient dT/dx. At the hot end of the rod both the concentration and the velocities of the electrons are higher than at the cold end. Therefore more electrons will start to diffuse in the direction of the temperature gradient than in the opposite direction. The diffusion flux, carrying the negative charge away from the hot end and transferring it to the cold end, sets up a potential difference between the ends.

The diffusion process is increasingly retarded by the electrical field in the interior of the semiconductor until the flux of electrons caused by diffusion is equal to the reverse flux caused by the potential difference which has arisen. In these conditions a dynamic electron equilibrium in the semiconductor will be established, under which the temperature difference between the ends of the rods will maintain a corresponding potential difference.

The number of electrons passing through any cross-section of the conductor in a unit time in both directions is equal. However, the velocities of electrons proceeding from the hot end are higher than the velocities of electrons passing through the given section from the cold end. This difference ensures that there is a continuous transfer of heat energy in the direction of the temperature gradient without any actual charge transfer.

The heat transfer mechanism is substantially different when negative (electrons) and positive (holes) charge carriers participate simultaneously in the current. Simultaneous transfer of equal numbers of holes and electrons does not lead to an accumulation of the charge and an increase of the potential. Simultaneous diffusion of electrons and holes from the hot end to the cold end is caused not only by the difference in the carrier velocities but also by their concentration gradient.

In the case of such a bipolar diffusion a thermal emf can also arise due to the following two factors:

1. When the concentration of one type of carrier (e.g. negative) exceeds the concentration of the opposite type of carrier the flux of the first type of carrier will carry towards the cold end predominantly a charge which will retard their motion and, conversely, accelerate the carriers of the opposite sign until the fluxes of both carriers are equal; an electric field governed by the temperature gradient will thus be established.

2. The difference of charge carrier mobilities forms the second source of thermal emf's. The mobility u is related to the diffusion coefficient D by the universal relationship established by Einstein:

$$\frac{u}{D} = \frac{e}{kT}.$$

Under the effect of the concentration gradient set up by the temperature gradient, the carriers characterised by higher values of u and D would have moved forward had they not created a space charge on becoming separated from the carriers of opposite sign. The corresponding electric field $\mathscr{E}$ retards their motion and accelerates the slower carriers of the opposite sign. The electric field $\mathscr{E}$ equalises the velocities of both types of carrier enabling them to diffuse as a single body.

Therefore, even when thermal agitation creates an equal number of current carriers of each sign, their diffusion produces an electric field in the conductor depending on the difference in carrier mobilities. The expression for this electric field $\mathscr{E}$ will be similar to that for the diffusion of ions in an electrolyte, namely

$$\mathscr{E} = \mathscr{E}_0 \frac{u_+ - u_-}{u_+ + u_-},$$

where $\mathscr{E}_0$ is the electric field which would have existed in the presence of current carriers of one sign only, and u_+ and u_- are the mobilities of positive and negative charges.

The advantage of semiconductors when used as materials for thermoelements is due to their higher values of thermal emf. We shall try to explain the reasons for this property of semiconductors.

In metals the concentration of free charge carriers is high (of the order of 10^{22} cm^{-3}) and does not depend on temperature; the kinetic energy of the highly degenerate electrons is virtually independent of temperature, and the same applies to the contact potential of the metal at its boundaries. Under such conditions thermal emf's as a rule do not exceed a few μV per °C.

In semiconductors, however, temperature has a pronounced effect on the concentration n and kinetic energy ϵ of free charge carriers, whilst

the absolute value of the concentration is smaller by a few orders of magnitude (10^{14} to 10^{20} cm^{-3}). The contact potential ϕ of semiconductors with respect to metals and the corresponding chemical potential μ are also functions of temperature. Thermal emf's amount in this case to hundreds of μV per °C.

What part does each of the above enumerated factors (low concentration, and the temperature dependence of the concentration, kinetic energy and contact potential) play in the appearance of high thermal emf's?

In the volume of a semiconductor the temperature gradient sets up a diffusion flux due to the concentration gradient dn/dx and the gradient of the mean kinetic energy of free charge carriers $d\epsilon/dx$. When the charge carriers are of one sign, the diffusion flux which transports them sets up an electric field $\mathscr{E}$ counteracting the diffusion. In the equilibrium state the electrical current produced by the field $\mathscr{E}$ balances the diffusion flux.

We shall consider separately the phenomena occurring in the bulk of the semiconductor and at its boundaries.

A temperature gradient may produce a unidirectional diffusion flux in two ways: (1) because of a concentration gradient of free carriers $\partial n/\partial x$, when $n = f(T)$; and (2) because of variation of the coefficient of diffusion D with temperature. Variation of D with temperature may in turn stem from two causes: (a) D may depend on the temperature T of the medium in which the carriers move; and (b) D may vary depending on the velocity v or kinetic energy ϵ of electrons transported by diffusion.

It will readily be seen that thermal agitation in the medium and its temperature T cannot by themselves produce a unidirectional flow of charge carriers, whilst in each portion of the solid an equilibrium thermal agitation is established corresponding to temperature T_m. Therefore $\partial D/\partial T_m$ does not lead to the appearance of a thermal emf.

On the other hand, dependence of the coefficient of diffusion on the kinetic energy of electrons $\partial D/\partial \epsilon$ may give rise to a unidirectional flow of carriers, even at a constant concentration n, when in addition to the temperature gradient dT/dx there exists also a kinetic energy gradient $d\epsilon/dx$. In fact, when the coefficient of diffusion of the carriers (electrons and holes) increases with rising temperature, i.e. when $\partial D/\partial \epsilon > 0$, more carriers will pass through each cross-section in the direction from the hot end to the cold end than in the reverse direction. The carriers will accumulate at the cold end, setting up an electric field $\mathscr{E}_D$ which will balance the diffusion flux. Conversely, when $\partial D/\partial \epsilon < 0$, the diffusion flux directed from the cold end to the hot end will predominate; charge carriers will accumulate at the hot end and the sign of the field $\mathscr{E}_D$, and therefore also of the corresponding part of thermal emf α_D, will be opposite.

Therefore the equilibrium condition for carriers in an insulated specimen with free non-degenerate carriers of one sign may be expressed

in the form

$$\mathscr{E}un = D\frac{dn}{dx} + n\frac{dD}{dx} \tag{204}$$

where u is the mobility of free carriers.

1. We shall first consider a special case of a semiconductor in which $n = \text{const}$ and $D = f(T) = F(\epsilon)$. Let $\mathscr{E}_D$ be the electric field in this case and α_D the corresponding thermal emf.

The equilibrium condition becomes:

$$\mathscr{E}_D un = n\frac{dD}{dx} = n\frac{\partial D}{\partial \epsilon}\frac{d\epsilon}{dx}.$$

Using Einstein's relationship

$$\frac{u}{D} = \frac{e}{kT}$$

we obtain

$$\mathscr{E}_D \frac{dx}{d\epsilon} = \frac{kT}{eD}\frac{\partial D}{\partial \epsilon}.$$

In order to express the dependence of D on T we shall use the familiar properties of mobility u which is interrelated with D. The mobility of electrons is proportional to the time τ required by the electrons to cover the mean free path, and

$$\tau = \frac{l}{v} \propto \frac{l}{\sqrt{\epsilon}}$$

where l is the mean free path length.

The relationship between l and ϵ depends on the conditions of scattering of free charge carriers; it varies depending on the nature of the chemical binding forces, lattice defects, and impurities, and is usually expressed in the form

$$l \propto \epsilon^r.$$

We have therefore $\quad u(\epsilon) \propto \tau \propto \epsilon^{r-\frac{1}{2}}$

and since $\epsilon \propto T$, $\quad D \propto uT \propto \epsilon^{r+\frac{1}{2}}$

$$\frac{1}{D}\frac{\partial D}{\partial \epsilon} = \frac{r+\frac{1}{2}}{\epsilon}.$$

Substituting this value in the equilibrium condition, we obtain

$$\alpha_D = \mathscr{E}_D \frac{dx}{dT} = \frac{k}{e}(r+\tfrac{1}{2}).$$

2. Let us now put $D = \text{const}$ and $n = f(T)$; the corresponding component of thermal emf will be denoted by α_n. We then have

$$\mathscr{E}un = \frac{D\partial n}{\partial x} = D\frac{\partial n}{\partial T}\frac{\partial T}{\partial x}$$

$$\alpha_n = \mathscr{E}\frac{dx}{dT} = \frac{kT}{e}\frac{\partial \ln n}{\partial T}.$$

3. Finally, let us consider the effect of the temperature dependence of the contact potential ϕ on the thermal emf α_ϕ; this dependence is determined by the temperature dependence of the chemical potential μ since

$$\frac{\partial(\phi)}{\partial T} = -\frac{\partial \mu}{\partial T}$$

where

$$\mu = kT\ln\left\{\frac{nh^3}{2(2\pi m kT)^{\frac{3}{2}}}\right\}$$

$$\frac{\partial \mu}{\partial T} = -k\left(\frac{3}{2}+\frac{\mu}{kT}-T\frac{\partial \ln n}{\partial T}\right)$$

$$\alpha_\phi = \frac{\partial \phi}{\partial T} = +\frac{k}{e}\left(\frac{3}{2}+\frac{\mu}{kT}-T\frac{\partial \ln n}{\partial T}\right).$$

It will be seen that the expressions α_n and α_ϕ contain the term $T\partial \ln n/\partial T$ which is governed by the temperature dependence of concentration. However, in the expression for the total thermal emf

$$\alpha = \alpha_D + \alpha_n + \alpha_\phi = \frac{k}{e}\left(r+2+\frac{\mu}{kT}\right)$$

this term cancels out. The potential difference produced by the concentration gradient in the bulk of the semiconductors is balanced by the difference of the contact potential at its boundaries. The experimentally observed thermal emf is not related to the temperature dependence of the concentration of the current carriers $\partial n/\partial T$. On the other hand, the absolute value of the concentration n appears in the expression for α, through the chemical potential, in the form $\ln n$.

All this is confirmed by the example of semiconductors in which the concentration of electrons (small in comparison with that in metals) remains constant in a wide temperature range. Semiconductors of this type include, for example, lead sulphide and lead telluride containing an excess of lead. Despite the constant concentration, the thermal emf is of the order of hundreds of μV per °C, i.e. of the same order as in semiconductors with a pronounced temperature dependence of concentration.

The difference in electron velocities at the hot and cold ends of the semiconductor leads, however, to the appearance of a potential difference

between the two ends. The higher the electron concentration in the semiconductor, the lower is the electric field $\mathscr{E}$ required for transferring the same number of electrons as that which diffuses owing to the difference in the velocities at the hot and cold ends of the rod.

The thermal emf α may be regarded as an entropy flux of unit electrical charge. As we have seen, the value of α depends not only on the entropy difference between two substances or two portions of a single conductor at different temperatures, but also on the conditions of motion of the electrons. These conditions may, in turn, depend on the nature of the semiconductor, and on the mechanism of electron scattering in the transfer of electrons from one portion of the semiconductor to another. Therefore, the value of α is closely related to the mobility u which is governed by the same scattering mechanism.

On the other hand, the scattering mechanism, and therefore also the values of u and α in anisotropic crystals (with the exception of crystals with cubic symmetry), depend on the direction. Therefore, a thermal emf appears between two rods cut out from the same crystal in different crystallographic directions, and Peltier heat is generated when current is passed through such a couple. When the current lines within a homogeneous crystal are bent, Peltier heat is generated everywhere where the current lines have a curvature. This is the so-called Bridgman effect.

Calculation of the thermal e.m.f.

For metals, in particular monovalent metals, quantum theory yields the following relationship

$$\alpha = \frac{\pi^2 k^2 T}{3e}\left[\frac{1}{\mu} + \left(\frac{1}{l}\frac{dl}{d\epsilon}\right)\right]. \tag{205}$$

Here μ is the chemical potential of electrons, which in the case of metals is equal to the Fermi level energy, and l is the free path length of electrons with a kinetic energy ϵ. For metals one may assume that

$$l \propto \epsilon^2,$$

and since the electrons contributing to the current have ϵ approximately equal to μ, we have

$$\frac{1}{\mu} + \frac{1}{l}\frac{dl}{d\epsilon} = \frac{3}{\mu},$$

whence

$$\alpha = \frac{\pi^2 k^2 T}{e\mu}. \tag{206}$$

The value of μ is almost independent of temperature and therefore in monovalent metals α is proportional to the absolute temperature T:

$$\alpha \propto T. \tag{206a}$$

Expression (206) gives an order of magnitude of

$$\alpha \doteqdot 10{\cdot}86\frac{3\times 10^{-2}}{5} = 5{\cdot}2\mu V/°\mathrm{C}$$

which is in good agreement with experimental data.

An exception is lithium for which α is 41 $\mu V/°$C. Samoilovich and Serova have shown that in this case there takes place, in addition to the diffusion of electrons, a dragging of current carriers by thermal waves (phonons) from the hot to the cold end of the conductor. The possibility of dragging of electrons by phonons was first pointed out by Gurevich.

In semiconductors one has to take into account the possibility of the existence of both free electrons with a concentration n_- and holes with a concentration n_+. The thermal emf's produced by the electrons and the holes have opposite signs and therefore one is subtracted from the other. In this case the thermal emf is given by the following expression

$$\alpha = \frac{1}{\sigma T}[u_- n_-(2kT-\mu) - u_+ n_+(2kT-\mu+\Delta E_0)],$$

where σ is the electrical conductivity, u the mobility and ΔE_0 the width of the forbidden energy band.

Pisarenko derived an expression for α which is more convenient for computation purposes

$$\alpha = \frac{k}{\sigma}\left\{u_- n_-\left[A+\ln\frac{2(2\pi m_-^* kT)^{\frac{3}{2}}}{h^3 n_-}\right] - u_+ n_+\left[A+\ln\frac{2(2\pi m_+^* kT)^{\frac{3}{2}}}{h^3 n_+}\right]\right\}. \quad (207)$$

The value of the constant A depends on the electron scattering mechanism, and in particular on the value of $1/l.\,dl/d\epsilon$. The latter is in turn governed by the character of scattering centres.

For atomic lattices

$$l \propto \epsilon^0, \quad u \propto \frac{1}{T^{\frac{3}{2}}}, \quad A = 2. \quad (208)$$

For ionic lattices when $kT \gg h\nu_0$:

$$l = l_0\sqrt{\left(\frac{\epsilon}{h\nu_0}\right)}\exp\frac{h\nu_0}{kT}, \quad u \propto \exp\frac{h\nu_0}{kT}, \quad A = 2{\cdot}5. \quad (208a)$$

For ionic lattices when $kT \gg h\nu_0$:

$$l \propto \epsilon, \quad u \propto \frac{1}{T^{\frac{3}{2}}}, \quad A = 3. \quad (208b)$$

For scattering by vibrations of constant frequency

$$\left.\begin{aligned} & l \propto \epsilon^{-1}, \quad u \propto \frac{1}{T^{\frac{1}{2}}}, \quad A = 1. \\ \text{or} \quad & l \propto \epsilon^{-\frac{1}{2}}, \quad u \propto \frac{1}{T^{\frac{3}{2}}}, \quad A = 0{\cdot}5. \end{aligned}\right\} \tag{208c}$$

For scattering on impurity ions

$$l \propto \epsilon^{2}, \quad u \propto T^{\frac{1}{2}}, \quad A = 4. \tag{208d}$$

In general when $l \propto \epsilon^{r}$, $A = r+2$ and

$$u \propto T^{r-\frac{1}{2}} \tag{209}$$

when the scattering takes place on thermal fluctuations. When lattice defects which do not depend on temperature form the scattering centres, we have

$$u \propto T^{r-\frac{1}{2}}. \tag{209a}$$

In practical applications of thermoelectric phenomena, the simultaneous presence of holes and electrons is undesirable if the hole and electron mobilities u_+ and u_- do not differ much from each other. For these purposes it is also desirable to use semiconductors with a high electrical conductivity. Both requirements are satisfied by the introduction of impurities into the semiconductor, creating either free electrons or holes.

For a semiconductor with current carriers of one type only, expression (207) may be simplified and replaced by the formula

$$\alpha = \frac{k}{e}\left(A + \ln \frac{2(2\pi m^* kT)^{\frac{3}{2}}}{h^3 n}\right). \tag{210}$$

Expressions for α may also be written down in a different form. For atomic lattices with current carriers of one type

$$\alpha = \frac{k}{e}\left(2 - \frac{\mu}{kT}\right) \tag{210a}$$

or, denoting by ΔE the distance of the impurity level from the corresponding band edge, we have at $\Delta E \gg kT$

$$\alpha \fallingdotseq \frac{k}{e} \times \frac{\Delta E}{2kT}. \tag{210b}$$

The last expression requires that the value of α should decrease with increasing temperature proportionally to $1/T$ and tend to infinity as the temperature tends to absolute zero. This conclusion not only disagrees with experimental data, but is also at variance with the principles of

thermodynamics according to which not only $\alpha(T)_{T=0} = 0$, but also $(d\alpha/dT)_{T=0} = 0$.

This contradiction indicates that the approximations made in the derivation of expressions (210), (210a) and (210b) are inadequate and that some properties which manifest themselves at low temperatures have been neglected.

For ionic semiconductors

$$\alpha = \frac{k}{e}\left(\frac{5}{2} - \frac{\mu}{kT}\right). \tag{211}$$

The polaron model of ionic semiconductors evolved by Pekar leads to more complex functions $l(\epsilon)$ and therefore somewhat different formulae for α. Pekar's theory is found to be well substantiated in the case of ionic semiconductors, when the electrons diffuse simultaneously with the polarisation of the surrounding medium. For a constant temperature T according to Pekar one can put

$$l_{\text{pol}} \propto \epsilon^2.$$

For practical purposes it is convenient to write expression (210) as a relationship between α and σ, both of which can be readily measured. By introducing an impurity or an excess of one of the components of the compound it is possible to vary the concentration n within wide limits without appreciably affecting the mobility u. Therefore the expression for α may be rewritten in the form

$$\alpha = \frac{k}{e}\left[A + \ln\frac{2(2\pi m^* kT)^{\frac{3}{2}} eu}{h^3}\right] - \frac{k}{e}\ln(neu)$$

or

$$\alpha \fallingdotseq C - 86 \times 10^{-6} \ln\sigma \fallingdotseq C - 2 \times 10^{-4} \log\sigma. \tag{212}$$

There is a large group of semiconductors with high electrical conductivity in which the concentration n does not vary with temperature. The effective mass m^* is also independent of temperature. According to expression (210) the temperature dependence of α for these materials takes the form

$$\alpha = \frac{k}{e}\left[A + \ln\frac{2(2\pi m^* k)^{\frac{3}{2}}}{h^3 n}\right] + \frac{3k}{2e}\ln T. \tag{213}$$

In such cases expressions (212) and (213) are well substantiated by experiment. In the plot of α against $\log\sigma$ α is represented by a straight line with a slope of approximately 2×10^{-4}. The temperature dependence of α at a given constant concentration n also corresponds to a slow logarithmic rise of α with T.

With the increase of free electron concentration the system gradually changes into the degenerate state when it is no longer possible to apply

the limiting case of Boltzmann statistics in place of the Fermi-Dirac statistics, which in general governs the electrons.

These more general formulae yield for the concentration

$$n = \frac{4\pi(2m^* kT)^{\frac{3}{2}}}{h^3} F_{\frac{1}{2}}\left(\frac{\mu}{kT}\right).$$

The function $F_r(x)$ represents the Fermi integral taken over all the electrons with different kinetic energies ϵ and therefore over all values $x = \epsilon/kT$

$$F_r(x) = \int_0^\infty \frac{x^r\, dx}{\exp\left(x - \frac{\mu}{kT}\right) + 1}.$$

Fermi integrals cannot be expressed as analytical functions but detailed tables are available for a range of values of r so that their use does not present any difficulties.

The most general expression for α for a given current carrier concentration n of one sign has the form

$$\alpha = \pm\frac{k}{e}\left[\frac{r+2}{r+1}\cdot\frac{F_{r+1}\left(\frac{\mu}{kT}\right)}{F_r\left(\frac{\mu}{kT}\right)} - \frac{\mu}{kT}\right]. \tag{214}$$

In particular, for atomic lattices, in which $r = 0$

$$\alpha = \pm\frac{k}{e}\left[2\frac{F_1\left(\frac{\mu}{kT}\right)}{F_0\left(\frac{\mu}{kT}\right)} - \frac{\mu}{kT}\right]. \tag{215}$$

Kontorova has shown that Pisarenko's expression (210) is of such a form that it does not lead to large errors up to a value of about 1 for μ/kT despite the onset of degeneracy. The onset of degeneracy is assisted by a small effective mass and therefore a low radius of curvature of the $\epsilon(k)$ curve. Under these conditions the density of the levels decreases and relatively few electrons occupy the free levels up to a certain value of the energy, which then represents the Fermi level. This is observed, e.g., in InSb.

All the foregoing considerations and formulae are based on the assumption that in a temperature gradient it is possible to regard the state of each small portion of material, containing a large number of atoms and electrons, as an equilibrium state, each portion only being described by different parameters. Gurevich showed in 1945 that this is not always

true. The flux of phonons proceeding from the hot end to the cold end drags a larger number of electrons than that which follows from conditions of normal diffusion and an electric field. This factor may bring about an increase of the thermal emf.

In fact, soon afterwards, Serova and Samoilovich showed that the abnormally high thermal emf of lithium is due to the drag effect.

Elementary calculation of the possible drag effect in semiconductors, carried out in 1951 by Pikus, indicated that its influence on the thermal emf is negligible.

Experimental data for germanium and silicon have since shown, against expectations, that the phonon drag effect is very high at low temperatures.

The disagreement with Pikus's calculations is attributed to the fact that electrons do not interact with all the thermal vibration phonons but only a few phonons, the wavelengths of which are not less than the wavelengths of the electrons.

Phonons satisfying this requirement possess a much larger free path length and a much longer relaxation time than the bulk of the phonons in the solid. Moreover, these phonons possess the most marked directional properties along the temperature gradient. Therefore, the effect of dragging of electrons by such phonons is particularly large.

At extremely low temperatures the drag effect is small since the free path lengths of both phonons and electrons depend on the dimensions of the crystals from which the solid is composed. At higher temperatures only the long wave phonons still retain a free path length of the order of the crystal dimensions. At still higher temperatures, the free path length of all phonons is smaller than the external dimensions, but the free path length of long wave phonons exceeds by many times its average value which governs the thermal conductivity of the solid. With a still further rise of temperature this difference becomes less pronounced and at a sufficiently high temperature, for which Pikus's calculations are valid, it no longer has any effect.

Accordingly, the increase of thermal emf, above that given by the formulae listed above, becomes particularly pronounced in a certain temperature range—usually below room temperature—for germanium around 30° K and for silicon around 100° K; in the case of silicon α reaches the value of 50,000 $\mu V/°C$. A complete theory of these phenomena has been given by Herring.

A systematic study of thermal emf's of various semiconductors has been carried out by Gokhberg and Sominskii. Among the materials investigated by them there was only one substance, namely WO_3, which

satisfied equation (210b): in the temperature range from 90° K to 400° K the thermal emf was found to be inversely proportional to the absolute temperature.

In V_2O_5 the relationship given by equation (210b) was only observed above – 30° C (240° K); below this temperature α was proportional to the absolute temperature as shown in Fig. 122. It is remarkable, however, that despite the abrupt change in the temperature dependence of α the temperature dependence of resistivity does not undergo a noticeable change at 240° K. Log ρ varies linearly with $1/T$ in the entire range from liquid air temperature to 500° K.

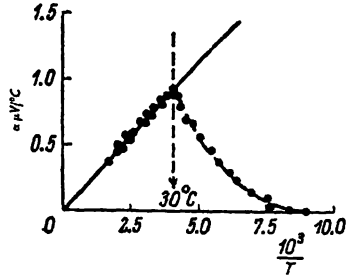


Fig. 122. Thermoelectric power of V_2O_5.

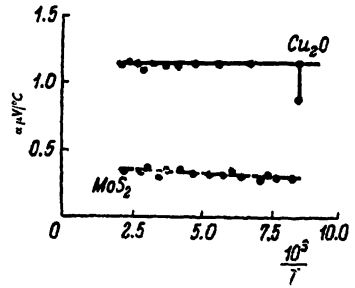


Fig. 123. Thermoelectric power of Cu_2O and MoS_2.

In Cu_2O α is found to be independent of temperature from 90° K to 630° K (Fig. 123) whilst above 630° K it is inversely proportional to temperature. Andersen has shown that in this case the sudden change of slope from a straight line for α against $1/T$ at 630° K coincides with a change of slope in the graph of $\log \rho =$ against $1/T$. The constancy of α below 630° K is at variance with theoretical predictions since here one would expect impurity conduction obeying expression (210b).

For tin sulphide (SnS) α is proportional to T at low temperatures; then α becomes constant, and at high temperatures it is proportional to $1/T$.

The temperature dependence of α is even more complex in the case of Tl_2S: at low temperatures α increases with an increase of T, then it decreases, again increases, and at high temperatures decreases once more.

The aforementioned semiconductors possess relatively high resistivities and a wide forbidden band. Data are also available for several other semiconductors of this type.

Expression (210b), according to which α is proportional to $1/T$, is found to be valid only at elevated temperatures (above 600° K in Cu_2O above 700° K in BaO, and above 700° K in NiO). On the other hand, at low temperatures α is proportional to T, which is contrary to expression (210b) but in full agreement with the requirements of quantum thermodynamics. Such a temperature dependence of α was found to hold for SiC below 200° K, CdO below 700° K, SbS below 190° K, and Ge below 30° K, and also in SnS, TiO_2 and certain other materials.

At intermediate temperatures there is often a range over which α is nearly constant; as has been mentioned earlier, this is true of Cu_2O in the range 90° K to 630° K. α was found to be constant in MoS_2 from 300° K to 500° K, in Bi_2S_3 from 110° K to 450° K, in SiC from 100° K to 300° K, and in PbS from 30° K to 750° K. In this temperature range the two tendencies—α proportional to T at low temperatures and α proportional to $1/T$ at high temperatures—appear to be balancing each other.

Modern theory of thermoelectric phenomena indicates that the thermal emf in impurity semiconductors with a non-degenerate system of electrons can only decrease with rise of temperature. According to this theory α can be proportional to temperature only in substances with degenerate electrons. In two cases (Ge and CdO), when such a temperature dependence of α was observed in semiconductors it was possible to demonstrate that the electrons were, as in metals, in a degenerate state.

There is, however, no reason to expect that in all semiconductors the electrons become degenerate as the absolute zero is approached. Incidentally, close to the absolute zero α and even $d\alpha/dT$ should tend to zero, since when T approaches 0° K, the entropy, according to Nernst's theorem, tends to zero, and, therefore, α also tends to zero.

Experimental data are in good agreement with such an extrapolation of the function $\alpha = f(T)$, but so far no theoretical foundations for this behaviour are available. General empirical formulae describing thermoelectric properties of semiconductors over a wide temperature range are also non-existent. It is only possible to state a qualitative rule according to which the thermal emf α of a semiconductor starting with $\alpha = 0$ at $T = 0$ is at first proportional to T, then rises less steeply with T often remaining constant in a certain temperature range, and then it begins to decrease approaching the theoretical expression (210a).

In order to calculate α it is necessary to know the effective mass m^* of current carriers and the scattering law. The value of m^* may be estimated on the basis of several other properties of semiconductors: from the activation energy of impurities, the mobility, and magnetic and optical phenomena. The nature of the carrier scattering governs the temperature dependence of mobility. The values of α for specific semiconductors calculated on the basis of such information are found to be close to the results of direct measurements.

Thermoelectric generators[65]

Let us consider a thermoelement (Fig. 124), consisting of p-type and n-type semiconductor rods joined by a metallic bridge. An external resistance R is connected in the circuit across the cold ends of the thermoelement; this resistance acts as a load for the electrical energy generated. When the potential difference across the cold ends of the circuit is denoted by V, the power delivered by the thermoelement will be W equal to V^2/R, whilst the current J is V/R.

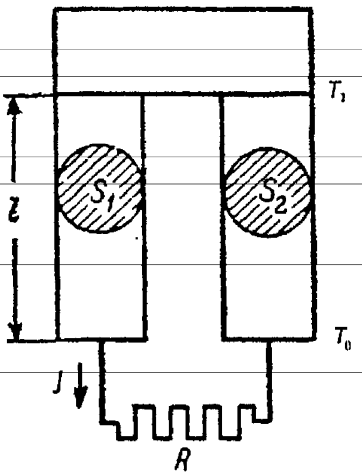


Fig. 124. Thermoelectric generator.

We shall denote the thermal emf, which is in this case equal to the difference between the thermal emf's of the two branches, by α.

The hot ends, joined by the bridge, are maintained at a temperature T_1, receiving heat energy from a source with a slightly higher temperature. The cold ends are at a temperature T_0, slightly above the temperature of the surroundings or of the heat exchange medium.

We shall denote the internal resistances of the two branches by r_1 and r_2 and their thermal conductances by K_1 and K_2. Denoting the resistivities by ρ, the thermal conductivities by κ, and assuming that the

lengths of both rods are equal to l, and their cross-section areas are S_1 and S_2, we have

$$r = r_1 + r_2 = \left(\frac{\rho_1}{S_1} + \frac{\rho_2}{S_2}\right) l,$$

$$K = K_1 + K_2 = \frac{(\kappa_1 S_1 + \kappa_2 S_2)}{l}.$$

On the basis of the general laws of thermoelectric phenomena we can calculate the Peltier heat generated and absorbed by the thermoelement at its two ends, the Thomson heat generated or absorbed inside the rods, the heat transferred by conduction from the hot to the cold ends, the Joule heat generated by the current in the rods, and the useful electrical energy delivered by the thermoelement. We shall refer to energy per second and express the power in watts. These units will also be used for purely thermal quantities; e.g. thermal conductivity will be expressed in W/deg cm and not in cal/deg cm sec.

(1 cal/deg cm sec = 4·19 W/deg cm.)

The amount of heat energy Q_1 received by the hot junctions, due to the Peltier effect, is

$$Q_1 = \alpha_1 J T_1.$$

The power Q_0 delivered at the cold junction is

$$Q_0 = -\alpha_0 J T_0.$$

The Thomson heat Q generated in each rod is, according to equation (198),

$$Q = \pm \int_{T_0}^{T_1} T \frac{d\alpha}{dT} J\, dT.$$

In the special case, when α at both ends has the same value, $Q = 0$.

The heat flux Q_h transferred from the hot junction to the cold junction through the two rods is

$$Q_h = K(T_1 - T_0).$$

The Joule heat generated in the two rods is $Q_J = J^2 r$. The useful power W delivered by the thermoelement is $J^2 R$; the current is

$$J = \frac{\alpha(T_1 - T_0)}{R + r}. \tag{216}$$

Putting

$$\frac{R}{r} = m,$$

we have

$$Q_1 = \alpha_1^2 T_1 (T_1 - T_0) \frac{1}{r(m+1)},$$

$$W = \alpha^2 (T_1 - T_0)^2 \frac{m}{r(m+1)^2}.$$

For the time being we shall neglect the Thomson heat, regarding it as small in comparison with the other terms and shall assume $\alpha_1 = \alpha_0 = \alpha$. When $\alpha_1 \neq \alpha_0$ the Thomson heat is accounted for in the determination of the efficiency by substituting for α the mean value for the two ends

$$\bar{\alpha} = \frac{\alpha_1 + \alpha_0}{2}.$$

Of the total Joule heat $J^2 r$ generated in the thermoelement, half passes to the hot junction, returning the power $1/2 . J^2 r$ and the rest is transferred to the cold junction.

The efficiency η will be defined as the ratio of the useful electrical energy $J^2 R$ delivered to the external circuit to the energy consumed from the heat source. The latter consists of the Peltier heat Q_1, and the heat Q_h transferred by conduction to the cold junction, from which it is necessary to deduct the electrical energy $1/2 . J^2 r$ returned to the heat source.

$$\eta = \frac{W}{Q_1 + Q_h - \frac{1}{2} J^2 r}$$

$$= \frac{\alpha^2 (T_1 - T_0)^2 \frac{m}{r(m+1)^2}}{\alpha^2 T_1 (T_1 - T_0) \frac{1}{r(m+1)} + K(T_1 - T_0) - \frac{1}{2} \frac{\alpha^2 (T_1 - T_0)^2}{r(m+1)^2}}$$

$$= \frac{T_1 - T_0}{T_1} \cdot \frac{\frac{m}{m+1}}{1 + \frac{Kr}{\alpha^2} \cdot \frac{m+1}{T_1} - \frac{1}{2}(T_1 - T_0) \frac{1}{m+1}}.$$

Thus the efficiency of the thermoelement is fully determined by (a) the hot and cold junction temperatures; (b) the quantity Kr/α^2 which depends on the properties of the materials used in the thermoelement and which will be denoted by $1/z$, so that

$$z = \frac{\alpha^2}{Kr} \tag{217}$$

and, finally, (c) the selected ratio $m = R/r$.

In order to achieve as high an efficiency as possible in thermoelements at given values of α, κ, and ρ and an arbitrary ratio m, it is necessary to find the optimum cross-section areas, S_1 and S_2 so that at given values of κ and ρ the product Kr is minimum. To find the condition for a minimum of Kr we shall differentiate

$$Kr = (\kappa_1 S_1 + \kappa_2 S_2)\left(\frac{\rho_1}{S_1} + \frac{\rho_2}{S_2}\right) = \kappa_1 \rho_1 + \kappa_2 \rho_2 + \kappa_1 \rho_2 \frac{S_1}{S_2} + \kappa_2 \rho_1 \frac{S_2}{S_1}$$

with respect to $d(S_1/S_2)$ and equate the derivative to zero. This gives

$$\frac{\rho_1 \kappa_2}{\kappa_1 \rho_2} = \left(\frac{S_1}{S_2}\right)^2. \tag{218}$$

At this value of S_1/S_2

$$Kr = [\sqrt{(\kappa_1 \rho_1)} + \sqrt{(\kappa_2 \rho_2)}]^2,$$

$$z = \frac{\alpha^2}{Kr} = \frac{\alpha^2}{[\sqrt{(\kappa_1 \rho_1)} + \sqrt{(\kappa_2 \rho_2)}]^2} \deg^{-1}. \tag{219}$$

This expression contains only the properties of the materials of the two branches of the thermoelement, but not their dimensions.

We shall now find the ratio m giving the highest efficiency. The condition for delivering the maximum power to the load leads, in the case of a thermoelement as for other current sources, to the requirement $R = r$, i.e. $m = 1$, and

$$\eta = \frac{1}{2} \frac{T_1 - T_0}{T_1 + \frac{2}{z} - \frac{1}{4}(T_1 - T_0)}.$$

We shall find the condition for the maximum efficiency by putting $d\eta/dm = 0$; straightforward calculations give

$$\left(\frac{R}{r}\right)_{\text{opt}} = M = \sqrt{[1 + \tfrac{1}{2} z(T_1 + T_0)]}.$$

Substituting this optimum value of m, denoted by M, into the expression for η we find

$$\eta = \frac{T_1 - T_0}{T_1} \cdot \frac{M - 1}{M + \frac{T_0}{T_1}}. \tag{220}$$

The first factor represents the thermodynamic efficiency of a reversible engine, and the second describes the reduction of this efficiency as a result of irreversible losses due to heat conduction and Joule heat entering into the expression for z.

The greater is M in comparison with unity, i.e. the larger the values of z and $(T_1 + T_0)$, the smaller is the reduction of the efficiency due to irreversible losses. Therefore an increase of the hot junction temperature T_1 increases η not only by increasing the value of the efficiency of a reversible engine $(T_1 - T_0)/T_1$ but also because of the simultaneous increase of M at a given z.

This equation also shows clearly that in order to achieve the maximum efficiency the material has to satisfy only one condition, namely the maximum value of z compatible with the maximum temperature T_1 of the heat source, or more precisely the maximum product $z(T_1 + T_0)/2$ attainable for the given material.

Assuming $T_0 = 300°$ K, we show in Table 13 the values of the efficiency η and M as functions of z at different values of T_1 and in Fig. 125 the dependence of η on T_1 for different values of z.

TABLE 13

Generator Efficiency as a Function of Figure of Merit and Temperature. Cold junctions at 300° K

T_1 °K	400		500		600		700		800		1000	
$z \times 10^3$	M	η%	M	η%	M	η%	M	η%	M	η%	M	η%
0·5	1·085	1·15	1·095	2·25	1·111	3·45	1·12	4·4	1·13	5·4	1·15	7·2
0·75	1·13	1·5	1·14	3·2	1·155	4·7	1·17	6·1	1·19	7·6	1·22	10·0
1·0	1·162	2·13	1·185	4·15	1·20	5·9	1·225	7·8	1·25	9·6	1·285	12·5
1·25	1·20	2·65	1·225	5·0	1·25	7·2	1·275	9·2	1·30	11·1	1·35	15·0
1·5	1·24	3·0	1·27	5·8	1·295	8·2	1·325	10·5	1·35	12·7	1·40	16·5
1·75	1·27	3·35	1·30	6·3	1·325	9·2	1·37	11·8	1·40	14·0	1·46	18·0
2·0	1·31	3·75	1·345	7·1	1·375	10·1	1·42	13·0	1·45	15·5	1·52	20·0
2·5	1·38	4·4	1·41	8·2	1·46	11·7	1·50	14·8	1·55	17·5	1·62	22·6
3·0	1·43	4·9	1·48	9·2	1·53	13·0	1·85	16·5	1·63	19·5	1·72	25·0
4·0	1·55	6·0	1·62	11·1	1·675	15·5	1·73	19·2	1·88	23·0	1·90	28·5
5·0	1·66	6·8	1·73	12·5	1·80	17·2	1·87	21·5	1·94	25·0	2·08	32·0

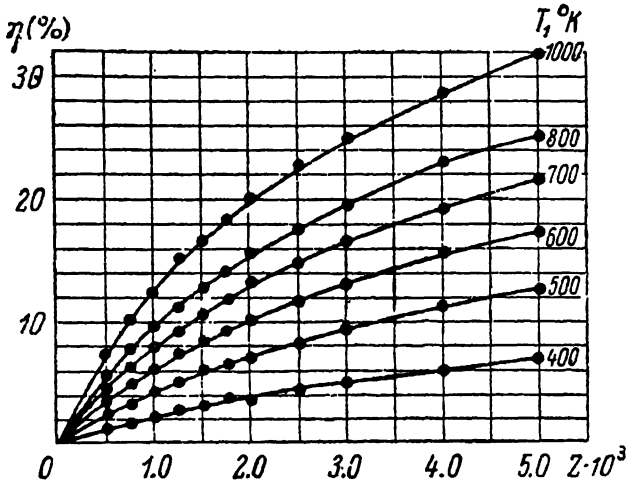


Fig. 125. Efficiency and figure of merit.

The efficiency was calculated from the expression

$$\eta = \frac{T_1 - T_0}{T_1} \frac{\sqrt{[1 + \frac{1}{2}z(T_1 + T_0)]} - 1}{\sqrt{[1 + \frac{1}{2}z(T_1 + T_0)]} + \frac{T_0}{T_1}} = \frac{T_1 - T_0}{T_1} \frac{M - 1}{M + \frac{T_0}{T_1}}.$$

Thermoelectric Cooling[66]

Thermoelectric refrigerators utilise the Peltier effect consisting of the generation or absorption of heat at the junction between two conductors when a current is passed through a thermoelectric circuit. We shall choose the direction of the current J so that the heat is absorbed from the cooled volume at a temperature T and generated at a room temperature T_0. The power absorbed at the cold junction is then

$$Q = \alpha JT$$

and that generated at the hot junction

$$Q_0 = \alpha_0 JT_0.$$

The Joule heat generated in the thermoelement is

$$W = J^2 r$$

where r is the resistance of both branches of the thermoelement. Half of W proceeds to the hot junction and the other half to the cold junction.

The heat transferred from the hot junction to the cold junction by conduction is

$$Q_h = K(T_0 - T).$$

In addition it is necessary to take into account the power consumed by the current flowing against the potential difference between the hot and cold junctions,

$$Q_p = \alpha J(T_0 - T).$$

The coefficient of performance ζ is equal to the ratio of the power absorbed by the cold junction to the electrical power consumed by the thermoelement

$$\zeta = \frac{\alpha JT - \frac{1}{2}J^2 r - K(T_0 - T)}{J^2 r + \alpha(T_0 - T)J}.$$

As in the case of the thermoelectric generator we shall describe the thermal conductance of the thermoelement K and its resistance r by the specific parameters for the two branches κ_1 and κ_2, ρ_1 and ρ_2 and the dimensions of the branches S_1 and S_2, $l_1 = l_2 = l$. As in the case of the thermoelectric generator ζ is maximum when

$$\frac{S_1}{S_2} = \sqrt{\left(\frac{\rho_1 \kappa_2}{\rho_2 \kappa_1}\right)};$$

under these conditions

$$Kr = [\sqrt{(\kappa_1 \rho_1)} + \sqrt{(\kappa_2 \rho_2)}]^2.$$

In the expression for ζ we shall multiply the numerator and denominator by r and replace Jr by V. Dividing the numerator and

denominator by V we obtain

$$\zeta = \frac{\alpha T - \frac{1}{2}V - Kr(T_0 - T)\frac{1}{V}}{V + \alpha(T_0 - T)}.$$

We shall select the potential difference between the ends of the thermo-element so as to make ζ maximum. For this purpose we shall equate $\partial\zeta/\partial V$ to zero which gives

$$V_{opt}^2 - \frac{4Kr}{\alpha}\cdot\frac{T_0 - T}{T_0 + T}V_{opt} - 2Kr\frac{(T_0 - T)^2}{T_0 + T} = 0,$$

whence $$V_{opt} = \frac{2Kr}{\alpha}\cdot\frac{T_0 - T}{T_0 + T}\{1 + \sqrt{[1 + \tfrac{1}{2}z(T_0 + T)]}\}.$$

Again putting $$\sqrt{[1 + \tfrac{1}{2}z(T_0 + T)]} = M,$$

we obtain $$V_{opt} = \alpha(T_0 - T)\frac{1}{M - 1}$$

and the value of the coefficient of performance corresponding to this value of V

$$\zeta = \frac{T}{T_0 - T}\cdot\frac{M - \frac{T_0}{T}}{M + 1}. \tag{221}$$

The maximum temperature drop $(T_0 - T)_{max}$ attainable at a given value of M will be obtained by equating ζ to zero. In this case we have M equal to T_0/T and

$$(T_0 - T)_{max} = \tfrac{1}{2}zT^2; \tag{222}$$

then $V_{opt} = \alpha T$ and the power consumption $W = \alpha^2 T^2/r = KzT^2$.

The first multiple in the expression for ζ represents the thermodynamic efficiency of a reversible Carnot cycle, and the second multiple the reduction of efficiency due to the irreversible processes of thermal conduction and Joule heat generation.

It is easily seen from the expression

$$\frac{1}{\zeta} = \left(\frac{T_0}{T} - 1\right)\frac{M + 1}{M - \frac{T_0}{T}}$$

that the coefficient of performance ζ decreases with increasing T_0/T and, for a given T_0/T, with decreasing M, i.e. decreasing value of $z(T_0 + T)/2$.

Taking for example $z = 2 \times 10^{-3}$ deg^{-1}, $T_0 = 300°$ K, and

$$\Delta T = 30° \text{ K},$$

we have $M = 1{\cdot}225$, and $\zeta = 0{\cdot}64$.

Table 14 gives the dependence of the value of ζ_{opt} on z and the temperature difference $(T_0 - T)$. In addition to the value of $(T_0 - T)_{max}$, following from expressions (221) and (222), Table 14 shows the corresponding values for thermoelectric batteries consisting of two stages

TABLE 14

Refrigerator Coefficient of Performance as a Function of Figure of Merit and Temperature Difference. Hot junctions at 300° K. *The Maximum Temperature Difference and the C.O.P. for Two Stages are also shown*

$z \times 10^3$	1·0		1·5		2·0		2·5		3·0		5	
stages	I	II	I	II	I	II	I	II	I	II	I	$\frac{T}{T_0 - T}$
$(T_0 - T)_{max}$ / $T_0 - T$	33	60	45	80	56	93	65	106	72	117	100	
5	3·6	3·7	4·9	5·0	6·3	6·3	7·6	7·6	8·9	8·9	12·4	59
10	1·4	1·6	2·2	2·25	2·8	2·85	3·3	3·4	4·1	4·15	6·1	29
20	0·44	0·52	0·83	0·89	1·1	1·2	1·45	1·5	1·8	1·85	2·7	14
25	0·2	0·30	0·55	0·64	0·78	0·86	1·0	1·05	1·3	1·35	2·0	11
30	0·1	0·23	0·33	0·45	0·56	0·65	0·78	0·85	0·96	1·06	1·6	9
35	—	0·16	0·21	0·34	0·41	0·50	0·56	0·66	0·74	0·85	1·3	7·6
40	—	0·11	0·12	0·26	0·29	0·38	0·44	0·56	0·58	0·70	1·02	6·5
45	—	0·07	—	0·21	0·20	0·31	0·33	0·44	0·45	0·58	0·84	5·7
50	—	0·03	—	0·15	0·11	0·24	0·22	0·33	0·33	0·47	0·68	5

arranged in series. It may be shown that, when the coefficients of performance of the individual stages are ζ_1 and ζ_2, the overall coefficient of performance for the two stages in series will be

$$\zeta = \frac{1}{\frac{1}{\zeta_1} + \frac{1}{\zeta_2} + \frac{1}{\zeta_1 \zeta_2}}$$

and when $\zeta_1 = \zeta_2$

$$\zeta = \frac{\zeta_1}{2 + \frac{1}{\zeta_1}}.$$

Straightforward calculations show that the use of two stages increases appreciably the value of ζ only when it is less than one. It is in practice useless to increase the number of stages above two.

When, however, it is intended to achieve the maximum temperature drop without paying attention to energy consumption, each new stage will decrease the temperature by a value somewhat smaller than $T(Tz/2)$. When z is independent of temperature, the reduction of temperature is proportional to T^2, where T is the final temperature achieved by cooling.

Therefore successive stages will give smaller and smaller temperature drops. When for example $z = 2 \times 10^{-3}$ deg^{-1} and $T_0 = 300°$ K, the first stage will produce a temperature drop of 59°, the second somewhat less than 40°, the third less than 29°, the fourth less than 22°, etc. The power consumption steeply increases with the increasing number of stages. The maximum temperature drop which may be achieved with a single stage at $z = 2 \times 10^{-3}$ deg^{-1} and $T = 100°$ K is only 8·5°.

Thermoelectric heating

Whilst cooling one set of junctions, a thermoelectric battery raises the temperature of the other junctions. In contrast to direct Joule heating, the electrical energy in a thermoelement serves for transporting additional heat energy from the cold to the hot junctions.

The calculations for such a heating device differ from those for a refrigerator in that the half of the Joule heat which is transferred to the cold junctions reduces their coefficient of performance, whilst that transferred to the hot junctions in a heating device is added to the Peltier heat.

The coefficient of performance in this case is

$$\zeta = \frac{\alpha JT + \frac{1}{2}J^2 r - K(T - T_0)}{J^2 r + \alpha J(T - T_0)} = \frac{\alpha T + \frac{1}{2}V - Kr(T - T_0)\frac{1}{V}}{V + \alpha(T - T_0)}.$$

Its maximum is obtained from the condition $\partial\zeta/\partial V = 0$ which gives

$$V_{opt} = \frac{\alpha(T - T_0)}{M - 1},$$

where

$$M = \sqrt{[1 + \tfrac{1}{2}z(T_0 + T)]}.$$

At V_{opt}

$$\zeta = \frac{T}{T - T_0}\left(1 - 2\frac{M - 1}{zT}\right).$$

At $z = 2 \times 10^{-3}$ deg^{-1}, $T_0 = 270°$ K and $T - T_0 = 30°$ K, $M = 1{\cdot}255$, $zT = 0{\cdot}6$ we have $T/(T - T_0) = 10$ and $\zeta = 1{\cdot}5$.

The values of ζ_{opt} are listed in Table 15 against z and $(T - T_0)$ for $T_0 = 300°$ K. The values of ζ can be raised by using two stages. With three successive stages, each raising the temperature by $\Delta T/3$, the coefficient of performance is still higher than with two stages, each raising the temperature by $\Delta T/2$. Thus, in raising the temperature by 30° from 300° K to 330° K the coefficient of performance is

z deg^{-1}	ζ_{I}	ζ_{II}	ζ_{III}
1×10^{-3}	1·30	1·38	1·46
$1{\cdot}5 \times 10^{-3}$	1·40	1·47	1·55
2×10^{-3}	1·65	1·68	1·71

TABLE 15

Coefficient of Performance of One- and Two-Stage Thermoelectric Heaters as a Function of Figure of Merit and Temperature Difference. Cold Junctions at 300° K

$z \times 10^3$	1·0		1·5		2·0		2·5		3·0		4		5		10	
stages / $T-T_0$	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	$\frac{T}{T-T_0}$
5	4·9	—	5·8	—	7·0	—	8·3	—	9·7	—	11·6	—	13·6	—	21·0	61
10	3·1	—	3·5	—	4·0	—	4·6	—	5·2	—	6·2	—	7·3	—	10·6	31
15	2·1	—	2·3	—	2·7	—	3·1	—	3·7	—	4·4	—	5·0	—	7·3	21
20	1·8	1·8	2·1	2·1	2·3	2·3	2·6	—	3·0	—	3·6	—	4·0	—	5·6	16
25	1·6	1·65	1·75	1·8	2·0	2·05	2·3	—	2·5	—	3·1	—	3·5	—	4·7	13
30	1·3	1·38	1·4	1·47	1·65	1·68	1·8	1·8	2·15	2·15	2·5	2·5	2·8	2·8	4·0	11
40	1·0	1·25	1·2	1·32	1·45	1·5	1·6	1·7	1·8	1·8	2·1	2·1	2·3	2·3	3·2	8·5
50	1·0	1·18	1·05	1·22	1·25	1·34	1·45	1·5	1·6	1·6	1·8	1·85	2·0	2·05	2·7	7

References

1. A. R. Regel', "The relation between electronic conductivity of liquids and their structure" (Voprosy Teorii i Issledovaniya Poluprovodnikov i Protsessov Poluprovodnikovoi Metallurgii), *Izd. Akad. Nauk SSSR, Moscow*, 12–28 (1955).
2. A. L. G. Rees, *Chemistry of the Defect Solid State.* London, Methuen & Co., 1954.
3. "Defects in crystalline solids", Report of the Conference held at the H. H. Wills Phys. Lab., University of Bristol. London, 1955.
4. F. A. Kröger and H. J. Vink, "Relations between the concentrations of imperfections in crystalline solids", *Solid State Physics*, **3**, 310–435 (1956).
5. W. G. Pfann, "Principles of zone-melting", *J. of Metals*, **4**, 747 (1952).
6. R. N. Hall, "p–n junctions produced by growth rate variations", *Phys. Rev.*, **88**, 139 (1952).
7. M. Blackman, "The specific heat of solids", *Handbuch der Phys.*, edited by S. Flugge, Vol. III, Part 1, Crystal Physics I. Berlin, Springer Verlag, 325–382 (1955).
8. V. V. Tarasov, "Chain and lamellar structures in crystals and glasses and the quantum theory of thermal conductivity", *Trudy Inst. Kristallogr. Akad. Nauk SSSR*, 136–148 (1954).
9. G. I. Rekalova, "Some optical and photoelectric properties of orpiment", *Zhur. Tekh. Fiz.*, **22**, 143 (1952).
10. B. I. Davydov and L. M. Shmushkevich, "Theory of electronic semiconductors", *Uspekhi Fiz. Nauk*, **24**, 21 (1940).
11. W. Shockley, *Electrons and Holes in Semiconductors.* Van Nostrand Co., N. Y., 1950.
12. E. Conwell and V. F. Weisskopf, "Theory of impurity scattering in semiconductors", *Phys. Rev.*, **77**, 388 (1950).
13. R. Mansfield, "Impurity scattering in semiconductors", *Proc. Phys. Soc.*, **69B**, 78 (1956).

14. A. I. Ansel'm, "The effect of resonant carrier scattering by impurity centres on the electrical properties of atomic semiconductors", *Zhur. Eksper. Teor. Fiz.*, **24**, 83 (1954).
15. S. I. Pekar, *Issledovaniya po Elektronnoi Teorii Kristallov* (Study of the Electronic Theory of Crystals). Gostekhteoretizdat, Moscow-Leningrad, 1951.
16. A. F. Ioffe, "The status of our knowledge of semiconductors and outstanding problems", *Uspekhi Fiz. Nauk*, **60**, 173 (1956).
17. C. Erginsoy, "On the mechanism of impurity band conduction in semiconductors", *Phys. Rev.*, **80**, 1104 (1951).
18. E. J. W. Verwey, P. W. Haayman and F. C. Romeyn, "Controlled-valency semiconductors", *Philips Res. Rep.*, **5**, 173 (1950).
19. E. J. W. Verwey and E. L. Heilmann, "Physical properties and cation arrangement of oxides with spinel structures", I, II, *J. Chem. Phys.*, 1947, **15**, 174–187.
20. K. Shifrin, "On the theory of electric properties of well conducting semiconductors", *J. of Phys. USSR*, **8**, 242 (1944).
21. B. I. Davydov, "The theory of solid rectifiers", *Dokl. Akad. Nauk SSSR*, **20**, 283 (1938).
22. A. F. Ioffe, "Rectification at the boundary of two semiconductors", *Zhur. Tekh. Fiz.*, **18**, 1498 (1948).
23. A. V. Ioffe and A. F. Ioffe, "Semiconductors in strong electric fields", *J. of Phys. USSR*, **2**, 283 (1940).
24. D. I. Blokhintsev and B. I. Davydov, "On the theory of solid rectifiers", *Dokl. Akad. Nauk SSSR*, **21**, 22 (1938).
25. W. Schottky and E. Spenke, "Quantitative approach to the space charge and contact layer theory of crystal rectifiers", *Wiss. Veröff. a.d. Siemens-Werkes*, **18**, 225 (1939).
26. J. I. Frenkel and A. F. Ioffe, "On the electric and photoelectric properties of contact between a metal and a semiconductor", *Phys. Zs. SU*, **1**, 60 (1932).
27. H. A. Bethe, "Theory of high frequency rectification by silicon crystals", NDRC Report, No. 43–11, Oct. 29 (1942).
28. A. I. Grubanov, *Teoriya Vypryamlyayushchego Deistviya Poluprovodnikov* (*Theory of the Rectifying Action of Semiconductors*). Moscow, Gos. Izd. Tekh. Teor. Lit., 1956.
29. H. K. Henisch, *Metal Rectifiers*. Clarendon Press, Oxford, 1949.
30. A. V. Ioffe and A. F. Ioffe, "Relaxation processes in a rectifying semiconductor couple", *Zhur. Tekh. Fiz.*, **23**, 209 (1953).
31. H. C. Torrey and Ch. A. Whitmer, *Crystal Rectifiers*. N.Y.–London, 1948.
32. W. Shockley, *Electrons and Holes in Semiconductors*. N.Y.–London, 1950.
33. E. Spenke, *Elektronische Halbleiter: eine Einführung in die Physik der Gleichrichter und Transistoren*. Springen Verlag, Berlin, 1955.
34. A. V. Ioffe, "Electrical resistances at semiconductor–metal contacts", *Zhur. Eksper. Teor. Fiz.*, **15**, 721 (1945).
35. A. N. Arsen'eva, "External photoeffect in semiconductors", *Izv. Akad. Nauk SSSR, Ser. Fiz.*, **16**, 122 (1952).
36. A. A. Kubetskii, "Some results of realisation of the secondary electronic conversion principle", *Izv. Akad. Nauk SSSR, Ser. Fiz.*, **8**, 357 (1944).
37. S. V. Starodubtsev and V. I. Pavlov, "Secondary emission of semiconductors exposed to positive ion bombardment", *Zhur. Eksper. Teor. Fiz.*, **7**, 409 (1937).

38. O. Madelung, "Die Theorie der galvanomagnetischen Effekte in Halbleitern", *Halbleiter-probleme*, **2**, 76–94 (1955).
39. D. K. C. MacDonald and K. Sarginson, "Galvanomagnetic effects in conductors", *Rep. on Progress in Phys.*, **15**, 249–275 (1952).
40. G. E. Pikus, "Theory and galvanomagnetic effects in semiconductors taking carrier density into account", *Zhur. Tekh. Fiz.*, **26**, 22, 36 (1956).
41. K. B. Tolpygo, "Transport equations in semiconductor theory", *Trudy Inst. Fiz.*, No. 3, 52 (1952).
42. A. I. Ansel'm, "Carrier density distribution in a semiconductor specimen in the presence of the Hall effect", *Zhur. Tekh. Fiz.*, **22**, 1146 (1952).
43. L. S. Stil'bans, "Adiabatic and isothermal Hall effects in semiconductors", *Zhur. Tekh. Fiz.*, **22**, 77 (1952).
44. P. L. Kapitsa.
45. M. Justi, *Leitfahigkeit und Leitungsmechanismus fester Stoffe* (*Conductivity of and Conduction Mechanism in Solids*). Gottingen, 1948.
46. E. Justi, Report on galvanomagnetic experiments.
47. H. Weiss and H. Welker, "Transverse magnetic resistance variation in InSb", *Zs. f. Phys.*, **138**, 322 (1954).
48. I. V. Mochan, Yu. N. Obraztsov and T. V. Krylova, "Study of thermomagnetic effects in p-type germanium", *Zhur. Tekh. Fiz.*, **27**, 242 (1957).
49. F. G. Bass and I. M. Tsidil'kovskii, "The theory of Hall and Nernst–Ettingshausen effects in semiconductors with mixed conduction mechanism", *Zhur. Eksper. Teor. Fiz.*, **28**, 312 (1955).
50. E. V. Zavoiskii.
51. Ya. G. Dorfman, "Paramagnetic and diamagnetic conduction electron resonance", *Dokl. Akad. Nauk SSSR*, **81**, 5, 765–767 (1951).
52. A. F. Kip, "Experimental work on cyclotron resonance in semiconductors", *Physica*, **20**, 813 (1954).
53. A. Fairweather, F. F. Roberts and A. J. Welch, "Ferrites", *Rep. Prog. Phys.*, **15**, 142–173 (1952).
54. Yu. P. Maslakovets and Yu. A. Dunaev, "Measurement of the thermal conductivity of lead sulphide", *Zhur. Tekh. Fiz.*, **16**, 1101 (1946).
55. E. D. Devyatkova, "Study of the thermal conductivity of lead telluride", *Zhur. Tekh. Fiz.*, **27**, 461 (1957).
56. A. V. Ioffe and A. F. Ioffe, "Thermal conductivity of semiconductors", *Izv. Akad. Nauk SSSR, Ser. Fiz.*, **20**, 65 (1956).
57. R. Peierls, "Kinetic theory of thermal conduction in dielectric crystals", *Ann. d. Phys.*, **4**, 121 (1930); **5**, 244 (1930); **12**, 154 (1932).
58. I. L. Pomeranchuk, "Thermal conductivity of dielectrics above the Debye temperature", *Zhur. Eksper. Teor. Fiz.*, **11**, 246 (1941).
59. E. D. Devyatkova and L. S. Stil'bans, "Study of the effect of atomic impurities on lattice thermal conductivity", *Zhur. Tekh. Fiz.*, **22**, 968 (1952).
60. A. V. Ioffe and A. F. Ioffe, "Some rules which hold for the thermal conductivity of semiconductors", *Dokl. Akad. Nauk SSSR*, **97**, 821 (1954).
61. A. F. Ioffe, "On thermal conduction in semiconductors", *Nuovo Cimento* (Supp. No. 4), **3**, 702 (1956).
62. G. Busch and M. Schneider, "Heat conduction in semiconductors", *Physica*, **20**, 1084 (1954).
63. A. I. Ansel'm and Yu. A. Firsov, "Free path length of a non-localised exciton in an atomic crystal", *Zhur. Tekh. Fiz.*, **28**, 151 (1955).

64. A. F. Ioffe, *Poluprovodnikovye Termoelementy*. Moscow, 1956. Translated into English in *Semiconductor Thermoelements and Thermoelectric Cooling*. Infosearch, London, 1958.
65. A. F. Ioffe, "Thermoelectric generators", *Izv. Akad. Nauk SSSR, Ser. Fiz.*, **20**, 76 (1956).
66. A. F. Ioffe, L. S. Stil'bans, et al., *Termoelektricheskoe Okhlazhdenie*. Moscow, 1956. Translated into English in *Semiconductor Thermoelements and Thermoelectric Cooling*. Infosearch, London, 1958.

CHAPTER

6

Methods of Measuring the Basic Properties of Semiconductors

THE effects observed in semiconductors may be described in terms of:

1. the electrical conductivity σ;
2. the electron and hole mobilities, u_- and u_+ respectively, and the relaxation time $\bar{\tau}$;
3. the electron and hole concentrations, n_- and n_+ respectively, and the corresponding effective masses m_-^* and m_+^*;
4. the forbidden band width ΔE_0, the distance between the band edge and an impurity level ΔE, and the chemical potential μ;
5. the lifetime τ_0 and the diffusion length L;
6. the thermal e.m.f. coefficient α;
7. the thermal conductivity κ, the thermal capacity per unit volume c' and the velocity of sound v;
8. the contact potential V_c;
9. the photoconductivity σ_p;
10. the Hall and Nernst constants R and N;
11. the dielectric constant ϵ;
12. the magnetic susceptibility χ;
13. the optical absorption coefficient k and the refractive index n.

We shall now consider the methods of determining all these properties in any given semiconductor, and how the parameters may be adjusted as required.

We discuss first the structure of semiconductors; in most cases we have to deal with microcrystalline aggregates produced either directly by oxidizing a metal (Cu_2O), by sintering compressed powders, or on solidification from the melt. In all these cases the specimens are inhomogeneous, and the conditions, under which the additives (an excess or deficit of oxygen or sulphur, for instance) are introduced, differ as between the surface and underlying parts; the impurities are often localized in the surface layers of the crystals. The individual crystals usually differ from one another, sometimes even in the sign of the charge carriers.

Adjacent crystals in the aggregates have differing orientations; crystals of the lower symmetry classes are anisotropic, in particular in their electrical and thermal conductivities. The electrical and thermal gradients in such materials are, therefore, variable and not susceptible of any simple interpretation, so the mean σ and κ values which result have no simple physical meaning. In pressed materials there is often partial orientation of anisotropic crystals in the direction of compression and, hence, there is thermal and electrical anisotropy.

The crystal surfaces are sometimes coated with low-conductivity films; the measurements of σ may then give completely incorrect results, since the overall conductance is very low due to the insulating layers, in spite of a high conductivity within the crystals. Some workers have considered that the results obtained must relate to these films and not to the base material, but this is not always correct, since in many cases such films may actually be of increased conductivity.

These difficulties all disappear when single crystals are used, though it should not be forgotten that natural single crystals have impurities which greatly influence all their properties. The same applies to artificial single crystals, grown from the melt, which contain impurities introduced from the containing vessel or from the air, even for extremely pure starting materials. Single crystals containing less impurity than 1 part in 10^8 have recently been grown; impurities have then been added in known concentrations.

The deviations from stoichiometry and from thermodynamic equilibrium occurring in semiconductors must also be borne in mind. Metal oxides and sulphides vary in composition with the oxygen (or sulphur) pressure in the gaseous phase over the specimen. Excess of non-metal in specimens with hole conduction (Cu_2O, UO_2, Cu_2S) increases σ and alters all the other parameters accordingly, while a deficit of non-metal produces the same effect in electron conductors (ZnO, TiO_2, ZnS). The specimen and the vapour phase reach equilibrium relatively rapidly at high temperatures. The excess oxygen concentration N and the oxygen pressure p are related by the law of mass action; for Cu_2O we have $N \propto p^{\frac{1}{4}}$ and $\sigma \propto N^{\frac{1}{4}}$, so $\sigma \propto p^{\frac{1}{4}}$. For ZnO we have $\sigma \propto p^{-1/4\cdot3}$.

If an energy E_K is required to introduce an excess oxygen atom into Cu_2O, we have

$$N \propto p^{\frac{1}{4}} \exp(-E_K/kT) \quad \text{and} \quad \sigma \propto p^{\frac{1}{4}} \exp[-(\Delta E + 2E_K)/2kT].$$

This σ–p relationship is found to apply between 800 and 1000° C; at lower temperatures the oxygen diffuses so slowly that the deep-lying areas of the specimen remain unchanged for hours or days, so the conductivity is determined by the equilibrium established at high temperatures. The oxygen content within the specimen is the same

as if the material had been quenched. We must, therefore, distinguish semiconductors which are in equilibrium at high temperatures from those quenched to low temperatures. The dividing line between the equilibrium and quenched states is, of course, determined by the time for which the temperature has been maintained and by the nature of the material. The properties of a quenched semiconductor depend on the conditions obtaining at the temperature T_K from which it was quenched.

The carrier mobilities in the purest possible single crystals are often found to be an order of magnitude larger than in impure or polycrystalline materials. The large effects produced by changes of composition or method of preparation can be utilized in altering the properties in a desired direction; impurities can be added, or deviations from stoichiometry introduced, to vary σ, α and σ_p widely. The annealing time or quenching rate are also important.

The changes in the sign of the carriers produced by added impurities are used in making germanium diodes and transistors. Thin p-type layers are produced by treating the surface of n-type material, and vice versa; the required sharp boundaries between the p- and n-type regions are thereby produced. p–n junctions may be produced within a single crystal during growth or by diffusion. Selenium rectifiers are formed at high temperatures with layers of n-type cadmium selenide on the surface of the p-type Se.

The required barrier layer structure for good rectifying properties in cuprous oxide rectifiers is formed by heating the copper to 1050° C (to form Cu_2O), saturating with oxygen at 600° C, and then quenching to a low temperature.

Electrical conductivity measurements

The resistances at the contacts form the main source of error in measuring the electrical conductivity σ; insufficient area of contact is mainly responsible with low-resistivity material, while barrier layers at the contacts, arising from contact potential differences, are the cause with high-resistivity material. The effect in either case may be to reduce the current flowing by a factor 10 or more, so all such contact phenomena must be eliminated. This can be done by using electrode-free methods (where the currents are induced by alternating or rotating magnetic fields), or by measuring the potential difference V, between probes P_1 and P_2 (Fig. 126) placed on a section of length L, carrying a current J; the resistance R of the section is just V/J, and σ is given by

$$\sigma = (J/V)(L/S)$$

where S is the area of cross-section. Here the equipotential surfaces are

supposed to be planes normal to the axis; large errors arise if the specimen is inhomogeneous, since the current is determined by the mean conductivity over the whole cross-section, while the potentials are determined by the surface layers.

In neither of the above methods is the contact resistance between the crystal grains eliminated, so the material must either be prepared as a single crystal, or else measurements must be made at frequencies such that the capacitative component of the current through the surface layers is much larger than the resistive component.

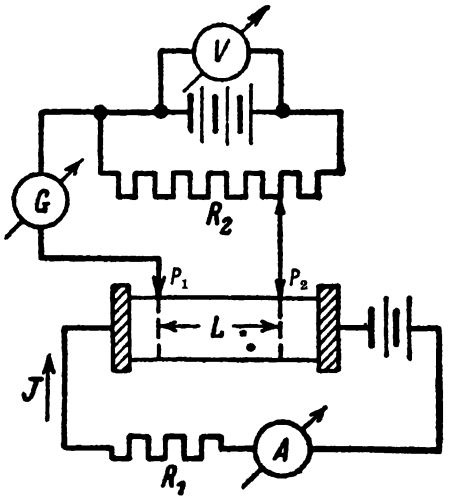


Fig. 126. Circuit for a null method of measuring electrical conductivities.

Guillery[1] suspended small semiconductor crystals in an insulating loss-free medium and thereby obtained a specimen with a high-frequency loss which was dependent on eddy currents within the crystals only; σ could therefore be determined. Insulating layers between grains are detected by raising the working frequency, since it is then found that the resistance becomes frequency-dependent. Kurchatov[2] and his co-workers have estimated the conductivity of silicon carbide crystals from the heating occurring in high-frequency fields. Regel'[3] has utilized the induced currents excited by a rotating magnetic field in measuring σ for molten semiconductors and for materials which react with metal electrodes at high temperatures. The specimen, suspended on a torsion fibre, twists round because the induced currents react with the exciting field. The steady deflection is a measure of the induced current, and hence of the conductivity.

Measurements on semiconductors of high resistivity, approaching the values to be found in insulators, involve very small currents; corrections must also be made for the back e.m.f.s produced by polarization. Currents in the range between 10^{-13} to 10^{-14} A can be measured in a few hundredths of a second with a fibre electrometer (Fig. 127), used as a null detector in a circuit containing the e.m.f. V (2–100 V), the resistance R to be measured and the balancing resistance R_0. This latter is joined to a potentiometer, which supplies a voltage V' such that E shows zero deflection. The current J through the specimen is then given by

$$J = V/R = V'/R_0,$$

$$R = R_0 V/V'.$$

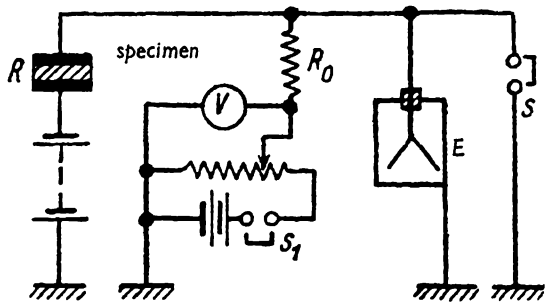


Fig. 127. Electrometer circuit for a null method of measuring high resistances.

Absence of a polarization e.m.f. must be checked when measuring R in this way; if J falls with V constant, or if the electrometer registers a charge on removing the e.m.f. V and the balancing voltage, polarization must be present. If the polarization voltage is V_n, R should be calculated from

$$R = R_0(V - V_n)/V'.$$

Still smaller currents, down to 10^{-16} A, can be measured with valve voltmeters.

Polarization effects can be avoided by working at mains or higher frequencies; any variation of resistance with frequency indicates low-conductivity barriers, or slow processes. Hysteresis effects are readily detected with an oscillograph.

Joule heating must be avoided when high voltages are used in resistance measurements, so short pulses or pulse trains are used, and the charge transferred per pulse, rather than the current, is measured.

In single crystals and orientated aggregates of anisotropic crystals σ varies with direction, so a complete description frequently requires that σ be measured as a function of direction.

Mobility

The mobility u may be determined from the speed of excess carriers injected from an electrode to which a short voltage pulse is applied.[4,5] The electric field $\mathscr{E}$ acting on the injected carriers and the velocity v of the leading edge of charge being known, u is calculated as $v/\mathscr{E}$. The increase in conductivity caused by the injected carriers alters the distribution of $\mathscr{E}$, but this effect is easily allowed for, as is the blurring of the leading edge caused by diffusion.

In single-crystal germanium the lifetime is about 0·001 sec, and the mobility is high, about 3000 cm^2/V sec, so the injected carriers can move some 20 cm in a 10V/cm field. A field of several hundred V/cm produced by a point can greatly increase the concentration of carriers of corresponding sign. The conditions required for the above method to be usable thus occur in germanium, and the results so obtained agree closely with those derived in other ways.[6] Another method of producing excess carriers is to illuminate the material for a short time with a narrow light beam, this producing the same effect as injection from a point. The mobility can be calculated from the time required for the carriers to reach a collector; this period is measured with an oscillograph.[7]

It is usual to measure u by combining measurements of σ and R, the Hall constant, which is (see Chapter 2) given by

$$R = A/nec$$

where n is the carrier concentration (if all carriers are of the same sign), e is the electronic charge, c is the velocity of light and A is a constant which depends on the law relating the mean free path l of an electron to its kinetic energy. In covalent lattices, with not too many impurity ions, A is $3\pi/8$; A is 1 in metals and ionic lattices at low temperatures, and 1·1 in ionic lattices at high temperatures. A is 1·93 if the ionized impurities are mainly responsible for electron scattering.

The product of R and σ, which is equal to neu, is Au/c, whence u is determined. The method is, however, only applicable to completely n- or p-type material, and is unsuitable for materials with mixed conduction. A semiconductor can often be made extrinsic by changing the temperature, or by introducing appropriate electropositive or electronegative impurities; u can then be measured. If the mobility so found is to be equated to that in the absence of added impurities, we must assume that the conditions for motion of the carriers have been altered only slightly. This assumption is only acceptable if the impurity concentration is low.

The Nernst effect can be utilized instead of the Hall effect, particularly in cases of mixed conduction; Mochan has obtained results, in this

way, in satisfactory agreement with those from other methods for pure germanium.[8]

The Hall and Nernst effects define the mobilities of the equilibrium carriers, while the drift method gives the excess carrier mobility. The data show that the mobility depends little on whether equilibrium or excess carriers are used.

R is difficult to measure when u and σ are small; it was for a long time impossible to determine the hole mobility in selenium, for instance, where u is less than 1 cm^2/V sec and ρ is 10^5 ohm cm. The accuracy of measurements of R can be greatly increased by using alternating fields and amplifiers, however.[9,10] The lowest mobilities observed (in Se, TiO_2, WO_3) are about 0·1 cm^2/V sec at room temperature, and the highest $3{\cdot}6\times10^3$ (Ge) and 1 to 8×10^4 cm^2/V sec (HgSe, InSb). The value for Ge rises to about 10^5 at low temperatures.

The equation,

$$u = el/2m^* v,$$

relates u to the mean free path l (the lifetime, of course, corresponds to many such free paths), while l is equal to $\bar{\tau}v$, where v is the thermal velocity.

Carrier concentration and effective mass

R gives the carrier concentration directly if the material is completely n- or p-type; in cases of mixed conduction u_+ and u_- have both to be determined. Now

$$\sigma = (n_+ u_+ + n_- u_-)e, \quad R = \frac{A(n_+ u_+^2 - n_- u_-^2)}{ce(n_+ u_+ + n_- u_-)^2},$$

so

$$R\sigma^2 = A(n_+ u_+^2 - n_- u_-^2)\, e/c.$$

n_+ and n_- can be derived separately, from σ and $R\sigma^2$ if u_+ and u_- are known.

The thermal e.m.f. α, which may be expressed in terms of

$$(n_+ u_+^2 - n_- u_-^2),$$

can also be utilized, though the other quantities appearing in the expression (particularly m_+^* and m_-^*) are not so accurately known as to give good values for n. The Hall coefficient has the important advantage of not depending on the effective mass, but is consequently of itself unsuitable for effective mass determinations, though it can be combined with α to give m^*. For a semiconductor with carriers of one sign only we have

$$\alpha = \frac{k}{e}\left[C + \ln\frac{2(2\pi m^* kT)^{\frac{3}{2}}}{h^3 n}\right].$$

Here C is determined by the law relating l to the velocity, as is A in the formula for R. C and A have values dependent on the type of scattering centre; C is equal to 2 in atomic lattices, to 3 in ionic lattices above the Debye temperature, and to 4 for scattering by charge centres. Hence m^* can be calculated from α, if n is found from the Hall effect, and the scattering law and thus A and C are known. As $\ln n$ appears in the formula, we do not need a very accurate value of n.

It is also possible to determine m^* from ΔE; if the dielectric constant $\epsilon > 10$ the impurity electron wave-functions extend over several lattice repeat periods and so the crystal structure can be assumed continuous. The ionization energy (the energy required to reach the free state) can then be taken as the value *in vacuo* divided by ϵ^2 and multiplied by m^*/m_0. For a monovalent hydrogen-like atom we have

$$\Delta E = \frac{13{\cdot}5}{\epsilon^2}\frac{m^*}{m_0}$$

whence m^* can be found.

A fourth method involves the measurement of n for the free carriers produced by impurities

$$n = \sqrt{(2N)}\left(\frac{2\pi m^* kT}{h^2}\right)^{\frac{3}{4}} \exp\left(-\frac{\Delta E}{2kT}\right)$$

where N is the impurity concentration.

A fifth method is the comparison of the mobility values for similar lattices using the formula $u = C'm^{*S}$, where S is close to $\frac{5}{2}$ for atomic and to $\frac{3}{2}$ for ionic lattices. However, C' also depends on m^*, so the method is unreliable.

m^* can be estimated from the variation in forbidden band width ΔE_0 with temperature.

Cyclotron resonance and changes of resistance in a magnetic field give the best and most complete data on effective masses; here the tensor properties of m^*, expressed in ellipsoidal form, become apparent. The orientation of the ellipsoid in the lattice and the longitudinal and transverse effective masses describe the complex shapes of the equal-energy surfaces in the Brillouin zone.

Cyclotron resonance data for Ge and Si have already been given. Diamagnetic or cyclotron resonances have not been demonstrated on all semiconductors, however, since the electron mean free path must be such that a complete orbit can be traversed without collision.

Forbidden band width and position of impurity levels

The temperature variations of σ and R provide the most useful method of measuring the forbidden band width and the position of an

impurity level. We have seen that σ is given approximately by

$$\sigma = \sigma_0 \exp\left(-\frac{\Delta E_0}{2kT}\right)$$

at fairly high temperatures if the impurity concentration is not too great. Now σ_0 contains a factor $T^{\frac{3}{2}}$, which can appreciably alter the shape of the σ curve at high temperatures; remembering also that $u \propto T^{-p}$, we have

$$\log \sigma = \log \sigma_0 + (\tfrac{3}{2} - p) \log T - 0{\cdot}43 \frac{\Delta E_0}{2kT}.$$

To a first approximation we neglect $(\frac{3}{2}-p)\log T$, and consider it only as a correction. The slope of the straight line obtained by plotting $\log \sigma$ against $1/T$ gives $\Delta E_0/2k$, and hence ΔE_0, when the above correction is applied. $\log \sigma$ could be replaced by $\log \sigma + (p - \frac{3}{2}) \log T$ in the plot.

ΔE can be derived better from the temperature dependence of R, R being mobility-independent. At low enough temperatures one sign of carrier (determined by the type of impurity) predominates completely, so

$$R = \frac{A}{nec} = \frac{A}{ec} \frac{h^3}{2(2\pi m^* kT)^{\frac{3}{2}}} \exp\left(\frac{\Delta E}{2kT}\right).$$

Assuming m^* to be temperature-independent, we get a linear variation of $\log R$ with $1/T$ (or strictly $\log RT^{\frac{3}{2}}$ and $1/T$).

The use of the temperature variations of σ and R has the advantage that the electron concentration is independent of the selection rules, which may forbid direct optical transitions, if thermal equilibrium is established, since an electron can always pass into a nearby free level. Spectral absorption or photoelectric methods do not have this advantage, as the nearest free levels may be inaccessible because of the selection rules, e.g. if the closest band consists of s levels only, no s electron can be transferred to it by a photon. The same selection rule applies to the wave vector k in a crystal lattice, so such transitions will be absent.

Thermal and optical transitions also have the important difference that in the latter the transfer is very rapid, the time required being about 10^{-15} sec (the periodic time of the light-wave). Ions cannot be displaced from their equilibrium positions in so short a time, and the only effect which can reduce the field is electron displacement, which corresponds to the optical range of the dielectric constant, i.e. the square of the refractive index in the infrared. Ionic polarization is set up in the next 10^{-12}–10^{-13} sec, and the field gradually weakens. The first component of the dielectric constant is of magnitude 3 to 20 units in ordinary semiconductors, while the second, slowly varying, part (due

to ion displacement) is 10–100 units; the energy of the electron immediately after transfer is therefore larger than when the ionic polarization is established. The photon energy required to produce a free electron in an ionic crystal is, thus, somewhat larger than the distance between the ground state and the excited state in thermal equilibrium. Polaron theory deals with this effect.

The light can also transfer electrons to excited states lying near the free band; the electron remains bound to its original level and takes no part in conduction. The primary effect of light absorption may even be to give an exciton with $h\nu < \Delta E_0$. The exciton has to diffuse through the crystal and to receive energy from a free electron, or from thermal sources, sufficient for one component (electron or hole) to pass to the conduction band. The lowest frequency absorbed, $\nu_{\min}$, does, therefore, not necessarily give the forbidden band width; in general, $h\nu_{\min} > \Delta E_0$ or $< \Delta E_0$; $h\nu_{\min} = \Delta E_0$ is only very approximate, though in many cases the agreement is found to be good.

ΔE_0 could be determined from the soft X-ray absorption and emission spectra (cf. Fig. 41), but the selection rules must be considered. If the K spectra (transitions to s levels) and L spectra (transitions to p levels) are recorded, account may be taken of the s, p and d electrons in the semiconductor. The X-ray methods are of low accuracy, however.

On the other hand, the temperature variations of σ and R give ΔE_0 correctly only if it is temperature-independent. If, as is usual,

$$\Delta E_0(T) = \Delta E_0(0) + \alpha T, \quad \ln \sigma = \ln \sigma_0 - \alpha - \Delta E_0(0)/2kT,$$

so the $\ln \sigma$ against $(1/T)$ plot will give a straight line of slope $\Delta E_0(0)$, and the intercept on the $\ln \sigma$ axis will be $\ln \sigma_0 - \alpha$, and not $\ln \sigma_0$. In many semiconductors α is 1 to 5×10^{-4} eV degree^{-1}.

The lifetime and the diffusion length

The lifetime and diffusion length are related to the diffusion constant D by

$$L = (D\tau_0)^{\frac{1}{2}},$$

while D is related to u by Einstein's universal law

$$D/u = kT/e.$$

Hence L can be expressed in terms of u

$$L = [(kT/e)u\tau_0]^{\frac{1}{2}}. \tag{224}$$

τ_0 may be measured at the same time as the drift mobility.[11] The recombination conditions are not altered by applying a weak extracting field between emitter and collector. The recombination rate and τ_0 can be determined by finding how the number of excess carriers, which are

collected, depends on the applied voltage, and hence on the time the carriers are in the field,

$$n = n_0 \exp(-t/\tau_0).$$

As in mobility measurements, illumination can be used instead of charge injection from an emitter.

Another method is the measurement of the rate at which the additional conductivity, induced by illumination or carrier injection, dies away

$$\sigma_p = \sigma_0 \exp(-t/\tau_0).$$

Free carriers combine at the surface as well as within the semiconductor; surface recombination is very sensitive to the state of the surface.[12] A surface that has been polished with a solid powder causes almost all carriers reaching it to recombine; there is also reason to think that the electrical state of the surface layer, and any adsorbed ions, may influence the recombination.[13]

The diffusion length L is even more sensitive to the presence of lattice defects than is the mobility u. Thus, heat treatment of germanium produces acceptor levels, but the mobility remains at the original value; L is only restored by prolonged annealing. L is measured in terms of the additional conductivity produced by constant illumination applied to a narrow strip. The illuminated area is a variable distance x from the collector; the law used is

$$n = n_0 \exp(-x/L).$$

L is, as would be expected, different for electrons and holes, and strongly dependent on the distance from the surface, where recombination is very rapid.

The thermoelectric power

The thermoelectric power or thermal e.m.f. α is determined from the potential difference V produced when the semiconductor is placed between two metal contacts with a known temperature difference $(T_1 - T_2)$. If current is allowed to flow the measured thermoelectric voltage is incorrect. V must be taken as the potential difference for zero current, and so must be measured potentiometrically; V can also be measured to a good approximation by using a galvanometer which has a resistance much larger than that of the specimen.

The thermoelectric voltage V depends on the properties of the two materials in contact, but the absolute thermal e.m.f.s for semiconductors are so much larger than for any pure metals that the result is little affected by changing the metal contacts. Almost any metal can be used to determine the sign of the thermal e.m.f., but exact values require a

metal with a thermal e.m.f. as close as possible to zero, such as lead or tin which also have very small Hall constants.

The thermal e.m.f. α is a function of temperature, so temperature differences of about 10° C should be used, with the mean temperature covering a wide range. In many semiconductor samples α changes in sign from part to part, so we must be able to measure at least the sign of α using very small volumes. A pointed copper rod, kept at a fixed temperature by means of an electric heater and brought into contact with the semiconductor, can be used. The other electrode remains at room temperature and is placed at some distant point on the specimen. The sign of the potential difference is then determined by the properties of the small area heated by the point.

The thermal e.m.f. α also depends strongly on the carrier concentration n and the electrical conductivity σ; α and σ are related by

$$\alpha = A - B \log \sigma.$$

α varies with n from a few tenths of a μV per ° C to 1000–1500 μV per ° C in certain semiconductors; α also depends on the scattering mechanism and on the effective mass m^*.

Thermal conductivity and heat capacity

Precision absolute measurements of κ are amongst the most lengthy and troublesome of operations, as the heat flux dQ/dt through an area S of the conductor and the temperature gradient dT/dx have to be measured

$$dQ/dt = \kappa S dT/dx. \tag{225}$$

The difficulty lies in determining dQ/dt, since some of the heat provided by the source (usually as Joule heat VJ, from a current J at a potential difference V) is lost to the surroundings by radiation or in other ways. It is difficult to apply corrections such that all errors are eliminated so methods in which the heat flux is not determined are often used. The heat is passed through the specimen and also through a standard substance of accurately measured conductivity κ_0 (Fig. 128). Neglecting heat lost to the surroundings, dQ/dt must be the same for both bodies. Then

$$\frac{\kappa}{\kappa_0} = \frac{S_0 \left(\frac{\partial T}{\partial x}\right)_0}{S \left(\frac{\partial T}{\partial x}\right)}.$$

The problem then amounts to determining the temperature gradients in the two bodies. The errors are reduced by mounting the apparatus in a vacuum and by heating the walls of the vessel to a temperature close to that of some appropriate part of the specimen.

A better method is to fasten the specimen between two rods of some material of accurately known conductivity. The heat fluxes $(\partial Q/\partial t)_1$ and $(\partial Q/\partial t)_2$ in these two rods are determined from the measured temperature gradients; the mean of the two is the flux in the specimen. Large errors are introduced by poor contacts between the bodies and the heat sources or sinks.

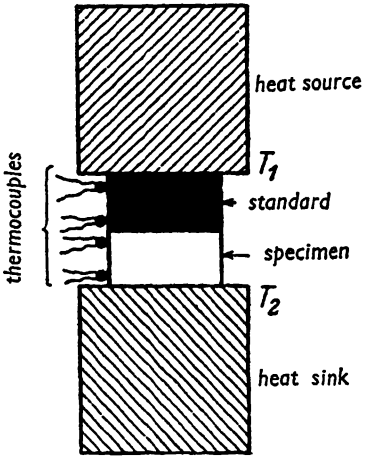


Fig. 128. A comparative method of measuring thermal conductivities.

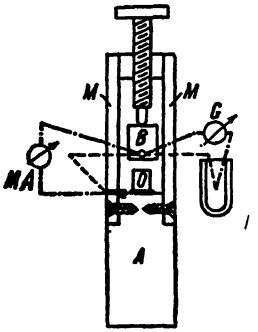


Fig. 129. A. V. and A. F. Ioffe's method of determining thermal conductivities. The specimen *O* is clamped between blocks *A* and *B*. *M* is an insulating jacket. The lead wires are of copper .-.-.-. and constantan – – –.

In all methods a steady state has to be set up, where the temperatures do not change with time. The time required may be some hours. Even the most careful measurements give an accuracy of only 1%, and failure to observe all necessary precautions may introduce large errors; results obtained by different observers on such standard materials as

rock salt and sylvine show large percentage errors. The steady-state method described above is used to measure semiconductor thermal conductivities over extensive temperature ranges.

Methods which do not require steady-state conditions are much more rapid. A. V. Ioffe and I have proposed one such method, for use with small specimens of semiconductors of thermal conductivities from 0·001 to 0·1 cal/cm sec ° C near room temperature.[14] The method is very simple and convenient, and will be described in some detail.

The principle is as follows: A small cylindrical specimen O, of low thermal conductivity, is clamped between two large copper blocks A and B (Fig. 129) at temperatures T_1 and T_2 which are measured by thermocouples. Heat is transferred between one block and the other (of thermal capacity c_B), via the specimen, when the first block is heated or cooled. The heat Q flowing through the specimen in time t alters the temperature T_2 by an amount ΔT_2. Neglecting lateral losses, we have

$$Q = c_B \Delta T_2.$$

Now for a specimen of cylindrical or rectilinear shape, of height L and area of cross-section S,

$$Q = \kappa(T_1 - T_2)\, St/L,$$

so

$$\kappa = c_B \Delta T_2 (L/St)/(T_1 - T_2). \tag{226}$$

The problem therefore amounts to measuring $(T_1 - T_2)$ and T_2 over a short period t.

The simplifications made in deriving equation (226) are as follows:

1. Thermal resistance at the contacts with the copper blocks is neglected. The error resulting from this cause is reduced by polishing the surfaces carefully and introducing a contact medium of conductivity higher than that of air (gallium, an amalgam, glycerine or oil). The error remaining is then small for materials of low conductivity (i.e. of about 10^{-2} cal/cm sec ° C).

2. The time required for the heat to spread evenly through the second block is neglected: it is assumed that, at all times, the temperature T_2 recorded by the thermocouple equals T_2', the temperature at the end of the specimen. It is also assumed that the galvanometer gives T_2 at the instant of measurement.

Of the quantities ΔT_2 and $(T_1 - T_2)$ appearing in equation (226), the former is not always affected by the lagging of T_2 behind T_2', if ΔT_2 remains the same for several successive intervals of length t. The error in $(T_1 - T_2)$ reduces κ somewhat, but $(T_1 - T_2) \gg (T_2 - T_2')$ normally, and the effect of replacing $(T_1 - T_2')$ by $(T_1 - T_2)$ is slight. The second block should not be of very large size in order to reduce this source of error; copper has a thermal conductivity hundreds of times

larger than has the specimen, and the time the copper takes to reach equilibrium is small, so the error is, in any case, slight. A simple correction factor of the form

$$1-\frac{\kappa L_0}{\kappa_0 L} \tag{227}$$

is sufficient to reduce the error considerably; here κ_0 is the conductivity of copper and L_0 the length of the copper block. The correction is usually about 1%.

3. T_1, T_2 and the temperature of the specimen vary continuously, so some of the heat flow, resulting from the temperature difference, is used to heat or cool the specimen. The temperature gradient in the specimen near the surface in contact with the top block will differ from the mean gradient $(T_1-T_2)/L$, but it is the former which determines the heat transfer to or from the top block. If the thermal capacity of the specimen is c_0 and that of the block c_B, the correction factor to be used is

$$1+\frac{1}{3}.\frac{c_0}{c_B}. \tag{228}$$

4. Heat loss to the surroundings, at a temperature T_0 (different from T_2), by radiation and conduction has been neglected; this error can be much reduced if the cylinder M round the top block and the specimen is kept at temperature T_2 during the measurements. A correction can also be readily applied. The heat lost by the top block can be put as $A(T_2-T_0)$; A is determined by allowing a steady state to be reached with the copper blocks at temperatures T_1' and T_2'. All the heat reaching the top block under the temperature difference $(T_1'-T_2')$ is then lost to the surroundings.

The interval of time t is selected such that ΔT_2 is readily measured, while (T_1-T_2) does not change too much. We have used t equal to 30 sec, with ΔT_2 about 1° C and (T_1-T_2) about 10° C.

Allowances must also be made for direct heat transfer from the bottom to the top block; the correction can be determined by replacing the specimen by another of known, very low conductivity, e.g. a hollow paper cylinder of the same size. The heat transferred, q, should be subtracted from Q. It is found that q, like Q, is proportional to (T_1-T_2), and inversely proportional to L,

$$q=(p/L)(T_1-T_2), \tag{229}$$

p is much larger in the case where the lower block is hotter than the upper and convection occurs, than when the lower block is the colder; in the first case p is $0{\cdot}8\times10^{-3}$, in the second $0{\cdot}3\times10^{-3}$ cal/cm sec ° C.

Applying all the above corrections, we have

$$\kappa = \frac{c_B + \frac{1}{3}c_0}{t} \cdot \frac{\Delta T_0 + \alpha(T_2 - T_0)}{T_1 - T_2} - \frac{p}{L}. \tag{230}$$

This formula gives values in error by not more than 3–5%, or sometimes even 1–2%, if $\kappa = 5\text{–}10 \times 10^{-3}$ cal/cm sec ° C; this is quite adequate for many purposes.

A fault in the method is that the heating rate must be measured over a finite time, so T_1 and T_2 change appreciably, while ΔT_2 is obtained as the difference of two readings. Zhuze and Regel'[15] have measured dT_2/dt simultaneously with T_1 and T_2 by passing the current from the T_2 thermocouple through the primary of a transformer with a permalloy core. The current flowing in the secondary gives dT_2/dt directly, which is either read simultaneously with $(T_1 - T_2)$, or passed on to a recorder; all data required are then available. In this way Zhuze and Regel' were able to measure conductivities very accurately by a dynamic method. The quantity actually measured directly was not κ, however, but the quantity K, which in the one-dimensional case is given by

$$\frac{dT}{dt} = K\frac{\partial T}{\partial x}S.$$

Comparing the expression for dT/dt and dQ/dt, and remembering that for 1 cm^3

$$dq = \rho c dt$$

where ρ is the density and c the specific heat, we have

$$\kappa = \rho c K. \tag{231}$$

In some cases we require κ, in others K; c and ρ are required to calculate one quantity from the other.

Zhuze and Regel's method is especially convenient for good conductors, where the temperatures change rapidly. It is better to record T_1 and T_2 at successive intervals if the conductivity is small.

The specific heat should be measured over a large temperature range; for this purpose some form of Nernst vacuum calorimeter may be used. The specimen, enclosed in a heating coil, is placed in a vacuum vessel, surrounded by liquid air if measurements are to be made at low temperatures, or by an oven for measurements above room temperature. A current is passed through the heater for a known time and the temperature rise recorded. Heat loss by radiation causes difficulties at high temperatures; the loss can be greatly reduced by silvering and polishing the specimen (if this is permissible) or the surfaces surrounding it.

The speed of sound is required to calculate the mean free path of phonons from the thermal conductivity; a piezoelectric ultrasonic

generator is used. The specimen is placed between emitter and detector; the phase difference produced by introduction of the specimen gives the speed of sound, while the change in amplitude gives the attenuation. Good acoustic contact at the surface is important. The speed of sound is not found to be frequency dependent in semiconductors, though some dependence on temperature and impurity content is observed.

The contact potential

The methods used for measuring contact potential are as follows.

1. In the mid-19th century Kelvin measured the difference of contact potentials between two parallel metal surfaces; a field $\mathscr{E}_c$ was produced in the space in between. The capacity of the assembly was varied by moving one plate relative to the other, and this altered the charge on the plates (Fig. 130); no change in potential occurred on displacing the plates. A potential V such as to make the field, and flow of charge, equal to zero, was applied to one plate, with the other plate insulated or connected to an electrometer. The applied voltage V was then equal in magnitude and opposite in sign to the contact potential V_c.[16]

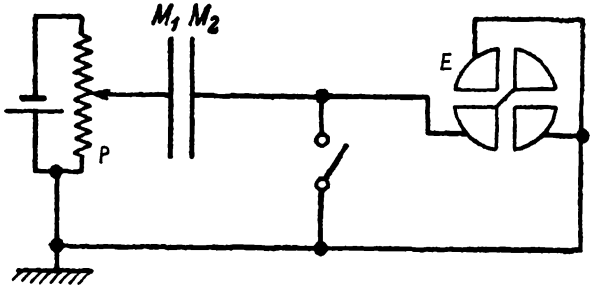


Fig. 130. Kelvin's capacitance method for determination of contact potential differences. M_1 and M_2 are the plates of the capacitor, E is an electrometer and P is a potentiometer.

Kelvin's method is insensitive, as large plates very close together, or very sensitive detectors, are required. With plates 1 mm apart, having a potential difference of 0·01 V, the charge per cm^2 is only 10^{-14} coulomb. The sensitivity is greatly increased by oscillating one of the plates at an audiofrequency and amplifying the resultant alternating signal.[17] The measurement is then accurate to 0·01 V, whether the metal or semiconductor be in air or vacuum.

2. The contact potential difference may be conveniently measured in ionized air; α-rays from polonium are used for the ionization, as these rays are rapidly absorbed. The ionizer is placed between the plates.

The contact potential difference is balanced out potentiometrically; the ionization current is reduced to zero, using an electrometer or electronic voltmeter to record current. Accuracies to 0·01 V are possible.

3. The longest wavelength of light which is capable of ejecting electrons from the body is often used to obtain the contact potential difference. The absorbed photons provide an energy $h\nu$, and the condition to be fulfilled is

$$h\nu \geqslant e\phi.$$

The minimum frequency ν_0 capable of exciting the external photoelectric effect, or the long wavelength limit (which is the same thing) corresponds to

$$h\nu_0 = e\phi.$$

The thermal vibrations must be taken into account when determining this limit, as they tend to give a blurred cut-off, though ϕ can be determined from the behaviour of the photocurrent near the threshold. This method gives satisfactory results with metals if proper allowance is made for the thermal vibrations and for the density of states near the Fermi level. The electron distribution between the energy states and the position of the chemical potential must be known for a semiconductor; even if we restrict ourselves to the electrons in the valence band, the threshold will not be as sharp as with metals.

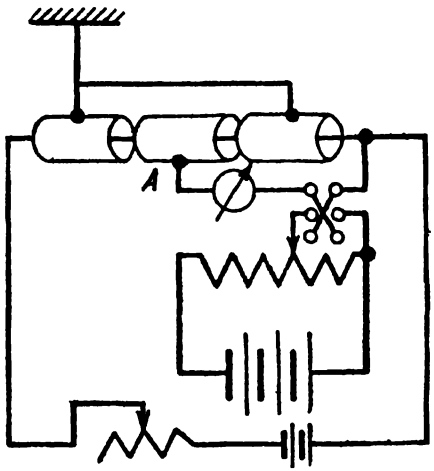


Fig. 131. Contact potential difference measurements using thermionic emission. A cylinder A, coated internally with the material, surrounds a tungsten wire.

4. Thermionic emission can be used to measure contact potentials (Fig. 131). At high temperatures the electrons are ejected with energies of about 0·1 eV, and are then accelerated or retarded by the contact

field produced by the other electrode. A potential difference sufficient to prevent the electrons reaching the second electrode is applied. These measurements are made under a very high vacuum; the emitter is a tungsten wire heated in the usual way and the test material forms a cylinder round the wire. The difference between the contact potentials of two substances may be measured by determining the potentials which have to be applied to retard the electron flow equally.

5. Attempts have been made to determine contact potentials from the energies of the electrons emitted from a heated wire into the space between it and another electrode.

Although the contact potential depends very much on the state of the surface, all the above methods give results in good agreement if the surfaces are identical. The optical method (photoelectric effect) may leave the semiconductor polarized, whereas the relatively slow removal of electrons in thermionic emission gives time for the polarization to die away. The photoelectric method may, therefore, yield a somewhat higher value for the work function of a semiconductor than do the thermionic emission or equilibrium contact potential measurements.

All the above methods, save only the photoelectric, give the work function for the outermost layers; the photoelectrons are ejected from a certain depth of material.

Photoconductivity

The photoconductivity σ_p can be defined as the difference between the electrical conductivities under illumination and in the dark; it should be related to the absorbed energy density or (more correctly) to the number of photons absorbed per cm^3. However, since reflectance and surface scattering measurements on the polycrystalline materials of which semiconductors are usually composed are difficult, it is usual to relate the increase in conductivity to the intensity and wavelength of the incident light. The intensity falls off inside the material; if the photocurrent and intensity are not linearly related, the photosensitivity may be given a wrong value, and the very sign of the photoeffect may change when the diffusion of photoelectrons is allowed for, since, although the primary effect of light absorption is to liberate fresh carriers, the conductivity may decrease on illumination (negative photoeffect). Non-uniform light absorption causes photoelectromotive forces, which themselves distort the observed change of conductivity.

The spectral dependence of the photoconductivity is also distorted because the fraction of the incident photons absorbed is a function of the wavelength and thickness of the material. When absorption is weak, much of it takes place at some depth, while strongly absorbed wavelengths are completely taken up in a thin surface layer and never

penetrate deeply. Curves relating the photoelectric spectrum to layer thickness such as those shown for Cu_2O in Fig. 132 are obtained.

The absorption coefficient k and refractive index n must be measured as functions of wavelength and the increased conductivity of any volume referred to the energy absorbed in it. We have shown previously that the photoconductivity is not necessarily proportional to the light intensity; a square-root variation is often found. The ratio of the photoconductivity to the absorbed energy is then a function of the light intensity. The photoconductivity depends on the concentration and nature of the impurities, and on the state of the surface for strongly absorbed wavelengths.

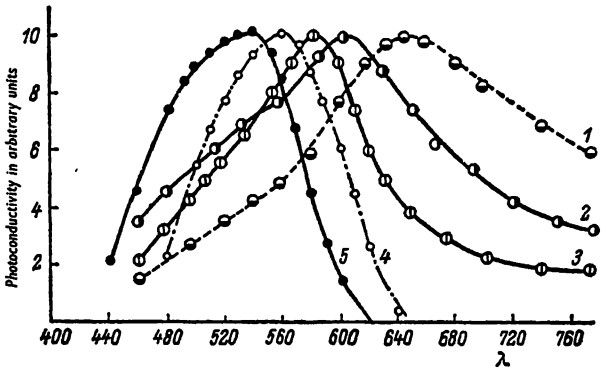


Fig. 132. Internal photoelectric effect in Cu_2O layers and the rectifier photoeffect in a cuprous oxide photocell (5). Layer thicknesses: (1) 300 μ; (2) 60 μ; (3) 20 μ; (4) 9 μ.

The photoelectric behaviour can be specified in terms of the number of carriers liberated per photon (the quantum yield), which is often not unity. Some semiconductors show no photoeffect at all; the light is absorbed and produces, say, excitons, of which but few dissociate to give conduction electrons.

The quantum yield must be determined in terms of the initial rise of photocurrent and not of the final equilibrium. Spark or gas discharges, or else a chopper disc inserted in a light beam, are used to provide brief excitation.[18,19] A Kerr cell is used to produce the shortest light pulses; a liquid showing double refraction in an electric field is placed between crossed polarizers, and light only passes when a field is applied to the liquid.

The lifetime τ_0 can be calculated from the leading edge of the photocurrent; the value is normally 10^{-5}–10^{-6} sec, but 10^{-3} sec is sometimes found. τ_0 can also be determined by illuminating the surface with a narrow light beam and measuring the rise in conductivity in adjacent

areas caused by carrier transfer (produced by an electric field or by diffusion). τ_0 can, finally, be determined from the number N of photoelectrons produced per sec, and the electron concentration n, which produces the photocurrent, for zero dark current

$$n = N\tau_0.$$

N can be determined from the initial rise in the photocurrent, and n from the Hall effect. The photocurrent differs from the dark current in that the light produces minority carriers as well as increasing the total carrier concentration.

The diffusion length L equal to $(D\tau_0)^{\frac{1}{2}}$ can also be determined by measuring the rate at which the current falls when the light is cut off; this is directly related to the electron lifetime.[20–23]

The dielectric constant

Although the dielectric constant ϵ is one of the most important parameters of a semiconductor, its relation to the distribution of states and to the impurities in the lattice has been little studied.[24] It is easily measured for semiconductors with $\rho > 10^7$ ohm cm, but for ρ between 10^3 and 10^5 ohm cm more difficulty is encountered; wavelengths in the centimetre or millimetre range are required.

There are three components of the dielectric constant, namely, (1) electronic, (2) ionic, and (3) orientation. The first can be calculated as the square of the refractive index in the long-wave infra-red region of the spectrum. The second can be determined by altering the equilibrium ion distribution, and the third by orientating the dipoles, which takes a time not greater than 10^{-10} to 10^{-11} sec. Electric field reductions often result also from charges moving through distances larger than the lattice constant and becoming trapped in lattice defects; these are really changes of the conduction mechanism.[25]

A fourth type of dielectric behaviour is found in ferroelectrics, where the dipoles are coupled together and show the same orientation over comparatively large areas. Here an external field rotates entire regions at once; adjacent regions show differing orientations. Adjacent dipoles may be coupled either parallel or antiparallel. Ferroelectricity, like ferromagnetism, only occurs within a specific temperature range; in both cases the effect vanishes above a critical temperature (the Curie point).

The electronic component of the dielectric constant can be determined from the refractive index, and the ionic and orientation components from radio-frequency measurements. The ferroelectric component is detected by extending the temperature range of the measurements. A ferroelectric semiconductor has a very high dielectric constant (some

thousands) which is highly temperature- and field-dependent; hysteresis effects are caused by residual polarization, and saturation occurs. These effects vanish above the Curie point.

The optical, magnetic and X-ray structural measurements made on semiconductors in no way differ from those for other materials. The magnetic properties are very closely related to the bond types present, and, hence, give valuable clues as to the structure of the semiconductor. New aspects of semiconductors have been revealed in studies using static magnetic fields, and high-frequency electro-magnetic fields at low temperatures. The soft X-ray emission and absorption spectra are important in determining the energy level structure, as are X-ray and electron diffraction data, since these latter give us the lattice structure and the time-average of the charge density, and, hence, the bond types. Studies of X-ray scattering by defects are important, and sometimes decisive, in our understanding of a semiconductor. These studies are, as yet, in their early stages, however.

One feature of semiconductor studies is notable: while there are thousands of papers dealing with the electrical and optical properties, there are only tens on thermal phenomena, and merely isolated unsystematic studies on the mechanical properties.

References

1. P. Guillery, "Conductivity measurement on powders", *Phys. Zs.*, **32**, 891 (1931); *Ann. d. Phys.*, **14**, 216 (1932).
2. I. V. Kurchatov, N. A. Kovalev, T. Z. Kostina and L. I. Rusinov, "Study of self-adjusting silicon carbide resistors", *Zhur. Tekh. Fiz.*, **3**, 1163 (1933).
3. A. R. Regel', "Measurement of the electrical conductivity of metals and alloys in a rotating magnetic field", *Zhur. Tekh. Fiz.*, **18**, 1511 (1948).
4. J. R. Haynes and W. Shockley, "Investigation of hole injection in transistor action", *Phys. Rev.*, **75**, 691 (1949).
5. R. Lawrence and A. F. Gibson, "The measurement of drift mobility in semiconductors", *Proc. Phys. Soc.* B, **65**, 994 (1952).
6. M. Prince, "Drift mobilities in semiconductors—I. Germanium", *Phys. Rev.*, **92**, 681 (1953).
7. S. M. Ryvkin and R. V. Khar'yuzov, "A method of determining the mobility of minority current carriers injected by light", *Zhur. Tekh. Fiz.*, **25**, 563 (1955).
8. I. V. Mochan, Yu. N. Obraztsov and T. V. Krylova, "Investigation of thermomagnetic effects in p-type germanium", *Zhur. Tekh. Fiz.*, **27**, 242 (1957).
9. B. R. Russel and C. F. Wahlig, "New method for the measurement of Hall coefficients", *Rev. Sci. Instr.*, **21**, 1028 (1950).
10. V. P. Zhuze and S. N. Nikolaev, "A method of measurement of the Hall effect in semiconductors", *Zhur. Tekh. Fiz.*, **23**, 913 (1953).

11. A. Many, "Measurement of minority carrier lifetime and contact injection ratio", *Proc. Phys. Soc.*, **67**, 9 (1954).
12. H. K. Henisch, W. Reynolds and P. M. Tipple, "Measurements of surface recombination in germanium", *Physica*, **20**, 1033 (1954).
13. R. H. Kingston, "Review on germanium surface phenomena", *J. Appl. Phys.*, **27**, 101 (1956).
14. A. V. Ioffe and A. F. Ioffe, "A simple method of measurement of thermal conductivity", *Zhur. Tekh. Fiz.*, **22**, 2005 (1952).
15. V. P. Zhuze and A. R. Regel', "A method for the rapid determination of the thermal diffusivity of solids and fluids", *Zhur. Tekh. Fiz.*, **22**, 1376 (1952).
16. W. E. Meyerhof and P. H. Miller, "A modified Kelvin method for measuring contact potential differences", *Rev. Sci. Instr.*, **17**, 15 (1946).
17. B. M. Tsarev, *Kontaktnaya raznost' potentsialov* (*Contact Potential Difference*). Gostekhteoretizd., Moscow, 1949.
18. S. M. Ryvkin, "Mechanical modulator for producing steep wavefront light pulses", *Zhur. Tekh. Fiz.*, **25**, 1471 (1955).
19. N. A. Tolstoi, "Precision tau-meter with a mechanical light modulator", *Dokl. Akad. Nauk SSSR*, **102**, 935 (1955).
20. V. E. Lashkarev, "Study of the kinetics of photoconductivity of semiconductors by the null method", *Izv. Akad. Nauk SSSR, Ser. Fiz.*, **14**, 199 (1950).
21. E. K. Putseiko, "Use of the condenser method for the study of the spectral distribution of photoelectric sensitivity in semiconductors", *Izv. Akad. Nauk SSSR, Ser. Fiz.*, **16**, 34 (1952).
22. V. P. Zhuze and S. M. Ryvkin, "Determination of the sign of photo-current carriers", *Dokl. Akad. Nauk SSSR*, **62**, 55 (1948).
23. N. A. Tolstoi and P. P. Feofilov, "A new method for the study of relaxation processes", *Zhur. Eksper. Teor. Fiz.*, **19**, 421 (1949).
24. H. Fröhlich, *Theory of Dielectrics*. Oxford, 1949.
25. A. F. Ioffe, *The Physics of Crystals*. McGraw-Hill Book Co., N.Y., 1928.

CHAPTER

7

Semiconducting Materials

The study of semiconducting materials

Out of the almost inexhaustible variety of semiconductors only a few groups have been studied—those with definite technical applications.

The first semiconducting material, which attracted the attention of physicists even at the end of the 19th century and in the first two decades of the 20th, was selenium. Photoconductors were prepared from it, and one of the first systems for transmitting pictures along wires was based on it. Since then some hundreds of papers on the properties of selenium have been published. Interest in selenium revived again when selenium rectifiers and photocells appeared, and the stream of papers devoted to selenium has not ceased even today.

At the end of the second decade of the present century cuprous oxide attracted much attention as a material for rectifiers of alternating current and, subsequently, for photocells. For a long time, right up to the beginning of the Second World War, cuprous oxide was the main subject of study in the semiconductor field, and it is difficult to estimate the number of papers devoted to it. The systematic study of cuprous oxide was carried on by both physicists and chemists, and it seems probable that the literature on it is not inferior in volume to that on selenium.

Before the war, in addition to these two semiconductors, there were also investigations of aluminium oxide, used for electrolytic condensers, and of uranium oxide and other materials used to make thermistors and time delay relays.

The Second World War posed two new problems in the solution of which semiconductors played a decisive role. These were radar and radiation detection in the far infra-red region.

The first of these problems led to the use of germanium and silicon as detectors and crystal amplifiers of radio oscillations. Long before the war, crystal detectors for receiving radio waves were known. Even 20 to 30 years ago Losev not only studied the physical nature of crystal detectors made from silicon carbide,[1] but also discovered the phenomena of luminescence in silicon carbide and of amplification in zincite.[2]

Silicon carbide was also employed in lightning arresters; it was studied by Kurchatov and Rusinov.[3]

A large amount of research was carried out in connection with radar, chiefly in the U.S.A., which elucidated in very detailed fashion the properties of germanium. With germanium as a model, numerous theoretical ideas were developed in the semiconductor field, and new physical phenomena were discovered. It so happened that in structure (regular crystal lattice of the diamond type), in chemical composition (a four-valent element) and in electrical properties (extremely high mobility and long free path for electrons), germanium was a particularly suitable object of study. Following germanium, and for the same reasons, the study of materials with crystal lattices analogous to the diamond type was commenced, for example, InSb, AlSb, GaSb, AlAs, etc.

Lead sulphide and telluride for the observation of infra-red radiation out to 5–7 microns and, subsequently, cadmium sulphide and diamond, for counting alpha-particles and corpuscular radiation generally, became widely used in the post-war period, and in this connection an appreciable literature devoted to this group of semiconducting materials has appeared.

Vacuum technology had been making use of semiconductors long before the war—barium oxide, thorium oxide, caesium sulphide, caesium antimonide—as sources of thermionic emission, secondary emission, and photoelectrons. However, the application and study of these materials over a number of years had not led to their semiconducting nature being observed. Only in recent years did Soviet scientists (Morgulis, Khlebnikov and others) and after them foreign workers (Wright) also, link the characteristics of these materials with the general theory of semiconductors.

To an appreciable degree this situation was repeated in the field of phosphors and fluorescent materials. Vavilov, Levshin, Adirovich, Antonov-Romanovskii and other Soviet physicists have, over the last decade, constructed a theory of luminescence starting from the study of semiconductors.[4]

In the thirties Kurchatov and Kobeko discovered in Seignette salt the electrical analogue of ferromagnetism to which was given the name "ferroelectricity".[5]

In the forties Vul[6] discovered a new group of materials with extremely high dielectric constants. These were the titanates of the elements of the second group and, in particular, of barium. Vul showed that the latter was ferroelectric. Barium titanate is widely used in industry not only in this country but also abroad. Smolenskii broadened still further the class of ferroelectric materials,[7] and Rzhanov and Vul

observed in ceramic ferroelectric materials the very important phenomenon of piezoelectricity, which has been put to use in acoustical apparatus.[8]

Finally, the possibility of preparing high-quality magnetic materials from semiconductors has been developed in recent years. In these materials iron ions are partially replaced by ions of other metals. Semiconducting ferromagnetic materials enable losses due to Foucault currents to be eliminated from magnetic apparatus.

Together with semiconducting ferromagnetic materials there have also been found anti-ferromagnetic materials. In these the domains are formed by electrons with opposed spins.

In all the above applications of semiconductors the materials used have specific resistivities within the limits $1–10^{15}$ ohm cm.

Only semiconductors of this kind were studied until recently, when attempts were made to use semiconductors for energy conversion. For this purpose semiconductors of low resistivity (thousandths or even ten-thousandths of an ohm cm) have made their appearance. These are alloys of metals, possessing properties similar to those of metals, but at the same time differing from metals by the presence in their energy spectrum of a forbidden energy band. Their carrier concentrations lie at the limit of the onset of degeneracy. The theory of such semiconductors was worked out by Shifrin.[9] Their physical characteristics were established by the researches of Zhuze, Boltaks, Devyatkova, Maslakovets, Stil'bans, Mochan, and myself, as well as other Soviet scientists.[10]

The appreciation of energy conversion problems directed attention to the thermal properties and, in particular, to the thermal conductivity of semiconductors, on which depends the efficiency of thermoelectric batteries. In this connection much importance has been attached to the study of solid solutions and multi-component systems.

New technical problems also present new technical requirements in materials. No previously known materials satisfy these requirements. Furthermore, the existing theory of semiconductors is inadequate; it does not enable one to predict accurately the conditions which determine the important properties of mobility, effective mass, and their temperature dependence. Existing theory, which works very well with such simple materials as silicon and germanium, leads to disagreement with experiment in the study of more complicated substances with different kinds of chemical bond—the compounds with which the new energy conversion problems are being solved. For this reason the range of semiconducting materials being studied by physicists and chemists is at present being rapidly extended. It would appear that the forms of the well-established band theory, which have their origin in the periodicity

of long-range order, are undergoing criticism, and are being reconstructed. The problem of liquid and amorphous semiconductors has arisen, and the problem of metallic alloys with borderline properties has acquired special significance fifty years after it was first pointed out by Kurnakov.[11]

The main task, however, is not just the study of semiconductors and their properties, but to gain such control of them that one may deliberately choose the optimum conditions for each field of technical application (and there are at least twenty of these). In order to do this one must establish the relations between the electrical, thermal, optical and magnetic properties of semiconductors and their chemical compositions and crystal lattices, and one must study more deeply the effect of different impurities and departures from stoichiometric composition.

Chemists and physicists are now presented with the urgent problem of producing together a new field of compounds with semiconducting properties, of studying them over a wide range of temperatures, and of creating a theory which generalizes the experimental data obtained with them. The solution of this problem will open up new vistas and will in turn, it would seem, lead to the task of producing new types of rectifiers, photocells, thermistors, transistors, etc.

The classification of semiconductors

In the preceding chapters we have considered the main ideas and conclusions arising from the band theory of semiconductors, which is based on the generalization of the quantum theory of metals. The band theory in this form is only applicable to a limited group of semiconductors. Experimental studies of other groups of semiconductors require substantial changes in the band theory and, in addition, the creation of a more general theory embracing amorphous semiconductors.

Returning to a consideration of the experimental data that have been accumulated, we shall classify the most important types of semiconductor which differ both in kind of chemical bonding and in crystal structure.[12]

The following types of semiconductor can be distinguished:

1. Semiconductors with the diamond type lattice, which is characterized by four directed valence bonds. To this group belong: diamond, silicon, germanium, grey tin, and compounds of elements equally displaced about the fourth group, such as mercury selenide and indium antimonide.

2. Elements of Group VI: sulphur, selenium, tellurium, which have a complex molecular lattice constructed from closed rings (sulphur), or chains of atoms (selenium, tellurium), which are comparatively weakly linked together, and which, probably, are partially dissociated.

3. Ionic crystals, NaCl, KCl, $NaNO_3$, and other polar compounds.
4. Sulphides, selenides, and tellurides.
5. Oxides and spinels.
6. Compounds of elements of the transition metal group, which have unfilled inner electron groups.
7. Alloys of metals termed "daltonide" and "berthollide" by Kurnakov.
8. Amorphous and liquid semiconductors, solutions of Na and K in ammonia.
9. Insulating organic crystals, glasses, and polymers.

In this chapter we shall present some of the, often incomplete, data available in the literature on the properties of the above types of semiconductor.

Atomic lattices with valence bonds[13–16]

A typical example of the elements of Group IV is diamond, in which each atom of carbon is bonded via four valence bonds which point towards the vertices of a tetrahedron formed by neighbouring atoms.

The lattice of diamond, as of other semiconductors of this type, has cubic symmetry. The thermal vibrations of the lattice are non-polar and mainly correspond to the acoustic branch (although there is also an optical branch). The scattering of the electron waves by these vibrations is weaker than in crystals in which there are, in addition, polar optical vibrations. In agreement with this, semiconductors with the diamond lattice possess high mobilities for electrons and holes.

The semiconducting properties of diamond have been studied in connection with its use as a counter of nuclear particles. Every alpha-particle or proton produces in a diamond a large number of ionizations, which sharply raise the extremely low conductivity of the unexcited crystal. The passage of each such particle is marked by a jump in the conductivity.[17]

However, with increase in the number of particles the sensitivity falls. The cause of this "fatigue" of the counter lies with the charges of the ions which remain in the bulk of the crystal, after the withdrawal of the electrons by the electric current. It is necessary to await the dissipation of this bulk charge by electrical conduction for the initial sensitivity of the counter to be restored. The restoration process can be accelerated by raising the conductivity in an appropriate fashion (e.g. by heating).[18]

The width of the forbidden band in diamond is 5·3 eV.

The mobility of electrons in pure crystals of diamond is 900 cm^2/V sec, and the lifetime of electrons about 4×10^{-8} sec. The lattice constant of diamond is 3·56 Å.[19]

A second crystal modification of carbon is graphite. This has a layer lattice with a large particle density in each layer and with a large separation between layers.

The distance between neighbouring atoms of carbon within the close-packed planes of crystalline graphite is 1·42 Å, while the separation between the planes is 3·35 Å. Atoms in neighbouring planes are arranged in checker-board fashion. The atoms of each layer are displaced relative to the normal to the plane, and only atoms of each third plane are positioned above those of the first plane considered. The unit cell of graphite, as with diamond and other diamond-like lattices, consists of two atoms. Accordingly, there are present in its thermal spectrum both optical and acoustic branches. It can be supposed that within the planes the atoms are bonded with valence and metallic bonds. The bond between the planes, however, is due to Van der Waals-type forces.

As a result of the different crystal structure and the different kind of chemical bonding, the physical properties of graphite are markedly different from those of diamond. X-ray spectra of graphite show that in it the bands due to the *d* valence electrons are overlapped by the *s* electron levels. The band of filled levels in the energy state distribution is not sharply separated from the empty band.

Instead of an energy gap there may be, in this case, only a sharp minimum in the density of states (see Fig. 133, ΔE_0 may tend to zero).

The study of single crystals of graphite has led to the conclusion that, within the close-packed planes, graphite behaves as if it were a two-dimensional metal with an unfilled band in which the electrons have an effective mass $m^* = m_0$. However, in the direction perpendicular to the planes, graphite is a semiconductor with a forbidden band of levels and a high effective mass.[20]

In crystals of boron nitride, which has an analogous structure to that of graphite, the distance between neighbouring atoms within the close-packed planes is 1·45 Å (instead of 1·42 Å in graphite) and the separation between layers is 3·30 Å (in graphite 3·35 Å). However, in contrast to graphite the bonding between boron and nitrogen atoms has a partially ionic character, and the arrangement of atoms in neighbouring planes is such that atoms of boron and nitrogen form chains normal to the plane of the layers. The energy gap in boron nitride is 4·6 eV, which makes boron nitride an insulator (Fig. 134).

Ever since silicon came to be used in high-frequency technology it has been studied in great detail.[13] The lattice constant of silicon is 5·42 Å, and the distance between two neighbouring atoms is 2·35 Å. The first Brillouin zone, which for diamond lattices has the shape of a

rhombic dodecahedron, holds up to 4 electrons per atom, that is all the valence electrons. In agreement with the experimental data presented previously, one can assert with confidence that silicon is a typical semiconductor with an energy gap of 1·12 eV.

Until recently, on account of the temperature variation of its resistance, silicon (actually of inadequate purity) was thought to be a metal, and the semiconductor-like temperature variation of resistivity found in certain measurements was ascribed to the influence of badly conducting layers of silicon oxide between the grains of the finely crystalline aggregate.

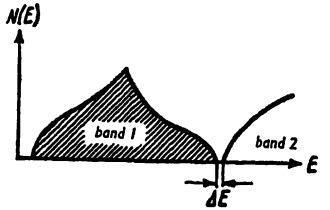


Fig. 133. Density of states distribution for graphite. The valence band (1) is separated from the next, conduction, band (2) by a very small energy gap (ΔE_0).

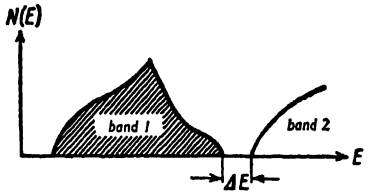


Fig. 134. Density of states distribution for boron nitride. The energy gap $\Delta E_0 = 4{\cdot}6$ eV.

The electrical conductivity σ of silicon has been found by various investigators to lie within the range 10^{-3} to 10^{3} ohm^{-1} cm^{-1}. The reason for this is a variation in the impurity concentration.

At the present time, as in the case of cuprous oxide, the intrinsic electrical conductivity of pure silicon, σ_1, and additional conductivity connected with impurity atoms, σ_2, are clearly distinguished. σ_1 and σ_2 are given by

$$\sigma_1 = A_1 \exp\left(-\frac{1{\cdot}12}{kT}\right), \quad \sigma_2 = A_2 \exp\left(-\frac{B}{2kT}\right)$$

where B varies from 0·06 eV for holes and 0·08 eV for electrons at low impurity concentrations, to extremely small values at large concentrations.

From these data it follows that the electrical conductivity of completely pure silicon at room temperature should be 5×10^{-6} ohm^{-1} cm^{-1}.

It can be supposed that impurity atoms, if their atomic volume does not differ markedly from that of silicon, replace silicon atoms in the crystal lattice. Silicon is found in the fourth column of Mendeleev's periodic system, and has four valence electrons. When an atom of silicon is replaced by an atom with a smaller number of valence electrons, for example by an atom of boron or aluminium (which are found in the third column), then such an impurity atom annexes an electron from the valence band of the crystal, leaving a free level. Accordingly, such atoms lead to hole conduction in silicon. By contrast, an element from the fifth column, such as phosphorus, having an excess electron which does not enter the valence band of the crystal, frees an electron which goes into the conduction band, leading to electronic conduction.

The energy required to free an electron, i.e. the work expended in transferring an electron to the conduction band, can be calculated in the following manner.

An electron going into an excited level leaves behind a positive ion with unit charge. The weaker the potential field V of the ion, the larger the radius of the orbit of the electron a. The potential field is determined by the dielectric constant ϵ of the medium

$$V = \frac{e}{\epsilon^2} \cdot \frac{1}{a} \qquad a = 0{\cdot}529 \times 10^{-8} \frac{m_0}{m^*}$$

where $0{\cdot}529 \times 10^{-8}$ cm is the radius of the orbit of an electron in a hydrogen atom, according to the Bohr theory.

The dielectric constant of silicon is 12·5, and the refractive index in the visible region of the spectrum is about 3·5 (which differs but little from $\sqrt{12{\cdot}5}$). Consequently, a is approximately $6{\cdot}6 m_0/m^*$ Å. If, for example, one assumes $m^*/m_0 = 0{\cdot}6$, we obtain $a_{Si} = 11$ Å. The dimensions of the orbit are larger than the interatomic spacing, so that one can consider the medium as continuous, and neglect its atomic structure.

The energy U of the electron in an orbit with principal quantum number n is

$$U = -\frac{e^4 m^*}{2\hbar^2 n^2} \cdot \frac{1}{\epsilon^2} = -\frac{13{\cdot}53}{\epsilon^2} \cdot \frac{m^*}{m_0} \cdot \frac{1}{n^2}.$$

For the first quantum orbit

$$U = \Delta E = -0{\cdot}087 \frac{m^*}{m_0} \text{ eV}.$$

Accordingly, the transfer of an electron from the first quantum orbit into a free state requires the expenditure of 0·087 eV on the assumption that for silicon m^*/m_0 is close to unity, or about 0·05 eV if $m^*/m_0 = 0{\cdot}6$.

The experimental value of the energy of activation for n-type silicon, with donor concentrations less than 10^{16} cm^{-3}, is about 0·075 eV, and is

almost independent of the chemical nature of the impurity atom. On increasing the concentration to 10^{19} cm^{-3} the value of ΔE decreases to zero. The difference between the calculated and measured values vanishes if one takes $m^*_{\perp}/m_0 = 0{\cdot}85$ (above we assumed $m^*_{\perp}/m_0 = 0{\cdot}6$).

The same kind of calculation of the energy required to free a hole gives a value slightly smaller than 0·06 eV, which is in good agreement with experiment if one takes $m^*_{\parallel}/m_0 = 0{\cdot}5$.

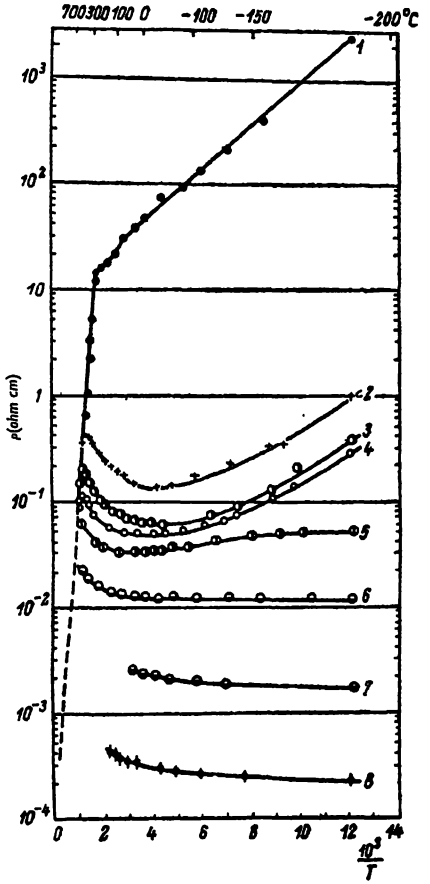


Fig. 135. The temperature dependence of the specific resistivity for silicon with small additions of boron (in weight per cent). From the data of Pearson and Bardeen. 1, pure; 2, 0·0005; 3, 0·001; 4, 0·002; 5, 0·005; 6, 0·01; 7, 0·1; 8, 1·0.

Measurements of the specific resistivity of the purest silicon, and for specimens of silicon with given concentrations of boron, are illustrated in Fig. 135. Measurements of the Hall constant are shown in Fig. 136.[21]

From the data in Figs. 135 and 136 the temperature dependence of the mobility of holes in silicon with additions of boron has been calculated (Fig. 137).

It is not difficult to appreciate the physical significance of these results.

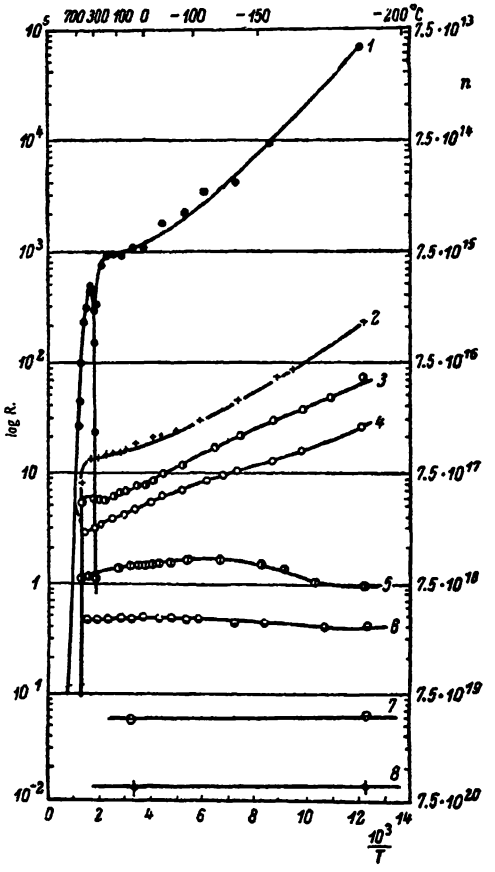


Fig. 136. The temperature dependence of the Hall constant and conduction electron concentration for silicon with small additions of boron. From the data of Pearson and Bardeen. Key as for Fig. 135.

In pure silicon at high temperatures, when the concentration of impurities is small, the concentrations of holes and electrons grow according to the law

$$n = A \exp\left(-\frac{\Delta E_0}{2kT}\right)$$

where ΔE_0 has the value 1·12 eV. Since the mobility of electrons is

greater than that of holes, the sign of the Hall effect corresponds in this region to electronic conduction. As the boron concentration is increased the number of holes increases also, becoming greater than the electron concentration, and changing the sign of the Hall effect from negative to positive. For small concentrations there is a temperature at which the Hall constant

$$R = \frac{3\pi}{8} \cdot \frac{n_+ u_+^2 - n_- u_-^2}{(n_+ u_+ + n_- u_-)^2} \cdot \frac{1}{ec}$$

passes through zero. The electrical conductivity σ, however, increases continuously with temperature since the marked increase in carrier concentration outweighs the effect of decreasing mobility.

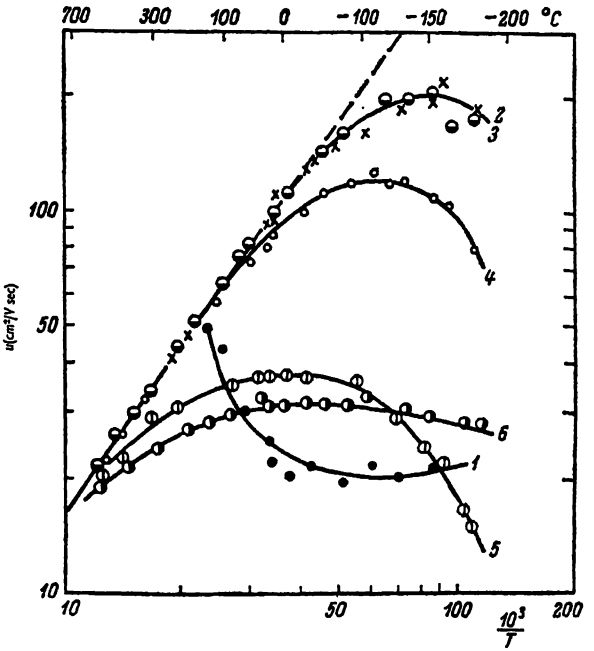


Fig. 137. The temperature dependence of the carrier (hole) mobility in pure silicon (1), and in silicon with small additions of boron (2–6, see Fig. 135). From the data of Pearson and Bardeen.

At low temperatures the concentration of holes is determined by the energy required to transfer electrons from the valence band to the atoms of boron. This energy, as obtained from the slope of plots of $\log \rho$ against $1000/T$ (Fig. 135), decreases with increasing boron concentration, and effectively reaches zero at a concentration of the order of $(5–6) \times 10^{18}$ cm^{-3}, when the wave functions of neighbouring impurity atoms overlap, and an impurity conduction band is produced.

The temperature variation of the mobility can be understood if one remembers that there act as scattering centres both thermal vibrations (which increase with temperature and lead to a law $u \propto T^{-\frac{3}{2}}$) and negatively charged boron ions (which give a dependence $u \propto T^{\frac{3}{2}}$).

The simultaneous action of both mechanisms of scattering should lead to a mobility u given by the expression

$$\frac{1}{u} = \alpha T^{\frac{3}{2}} + \beta T^{-\frac{3}{2}}.$$

At low temperatures scattering by the ions predominates. This decreases with rising temperature. At high temperatures the effect of thermal vibrations predominates, and this increases with temperature. Only curve 1 in Fig. 137, for which the impurity composition is unknown, is an exception to these rules.

On comparing Figs. 136 and 137 it is easily seen that the positive temperature coefficients of resistivity for the left hand parts of curves 2–6, which are anomalous for a semiconductor, are due to variations in mobility and not in the concentration of holes.

Accordingly we have, with silicon, a semiconductor for which all predictions of the theory are well fulfilled. In the diagrams shown, which accurately represent variations with temperature and the influence of impurities, the only quantitatively incorrect value indicated is the mobility of holes in silicon; as shown, it does not exceed 200 cm^2/V sec.

Direct measurements of the velocity of transport of charges in an electric field lead to higher values of the mobility, reaching 1200 cm^2/V sec for electrons and 250 cm^2/V sec for holes. The mobility of electrons in silicon, as might be expected, is larger than for diamond and smaller than for germanium.

Above 700° C silicon can be obtained in an amorphous modification, from which it crystallizes at lower temperatures.

Silicon is used in the manufacture of detectors for very short (centimetre and millimetre) radio wavelengths. A thin tungsten wire of diameter $d < 0{\cdot}1$ mm is used as one electrode. The resistance produced by such a contact on a semiconductor with a specific resistivity ρ is $\rho/4d$, and is mainly concentrated in a layer about 10^{-4} cm thick, close to the contact. To obtain a barrier layer the silicon is given a prolonged heat treatment which, apparently, produces an n-type layer on the surface of the p-type silicon. On account of the small capacitance of this kind of detector, and its good rectification factor, it can be used up to frequencies of 10^{10} cycles per second. Photocells based on p–n junctions have been prepared from very pure single crystal silicon in the U.S.A. The efficiency of such photocells, using solar radiation, i.e. the fraction of radiant energy converted directly into electricity, reaches 10%, which

is comparable with the efficiency of steam engines and exceeds by a factor of almost 100 the efficiency of photosynthesis in plants.[22, 23]

The electrical properties of germanium, the principal material for making high-frequency detectors and amplifiers, have been particularly carefully and thoroughly studied.[13, 16]

As is well known, the properties of germanium, which is an element of the fourth group, were predicted by Mendeleev long before its existence in nature was confirmed.

Crystals of germanium, as of diamond and silicon, belong to the regular system. They are formed by the valence bonds of a germanium atom with four nearest neighbours positioned at the apices of a tetrahedron surrounding it. The lattice constant is 5·647 Å, and the closest distance between atoms is 2·44 Å. Amorphous layers of germanium crystallize above 400° C. The energy gap is 0·78 eV, and the separation of impurity levels from the edge of the conduction band is about 0·015 eV, i.e. it is smaller than in silicon, which corresponds to the larger dielectric constant ($\epsilon = 16$), and the smaller effective mass $m^*_{\pm}$ of electrons in germanium ($m^*_{\pm}/m_0 \doteq 0{\cdot}3$).

The limit of strong optical absorption lies, for germanium, at a wavelength of 1·6 microns, which is in good agreement with the energy gap of 0·78 eV.

As in silicon, additions to germanium of elements of groups II and III (calcium, boron, copper) yield hole conduction, while additions of group V elements (bismuth, antimony, phosphorus, nitrogen) result in electronic conduction. Additions of tin, which also belongs to group IV, lead to electronic conduction.

On melting, germanium forms a liquid in which atoms of germanium are surrounded, not by four, but by eight neighbours, and in which the covalent bonds have been replaced by metallic bonds. Liquid germanium, as has been shown by Regel', is a metal.[24] The closer packing of the atoms leads to a decrease in the specific volume of germanium on melting.

The mobilities of electrons and holes in germanium are extremely high. Measurements of the Hall constant R and electrical conductivity σ lead to the same results as direct measurements of the velocity of charge transport by the drift method in an electric field (Fig. 138). This method has been used to measure the mobility u_- of electrons in p-type germanium, and that, u_+, of holes in n-type material. In pure single crystal germanium it appears that at room temperature $u_- = 3600$ cm²/V sec and $u_+ = 2700$ cm²/V sec.

The mobility of holes in n-type germanium has also been measured by their diffusion from an illuminated part of a crystal without any

external field being applied, and besides this, by the usual method of combining Hall effect and electrical conductivity measurements. All three methods lead to results in good agreement with each other.

As the temperature is lowered the mobility of electrons increases in inverse proportion to $T^{\frac{3}{2}}$ (more accurately to $T^{-1.65}$). At liquid nitrogen temperature (about 80° K) it reaches 30,000 cm²/V sec.

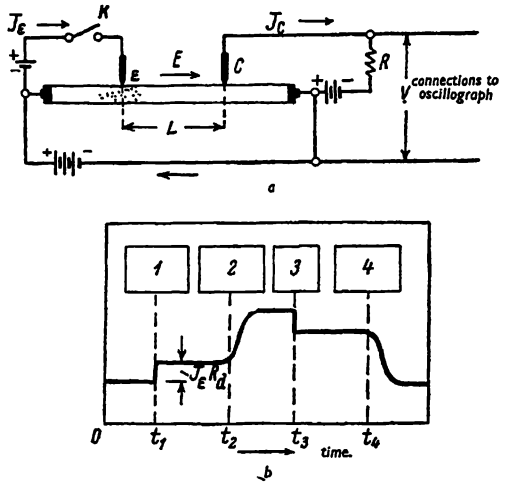


Fig. 138. Investigation of the motion of holes injected into *n*-type germanium: (a) experimental set-up; (b) form of signal on the oscillograph screen. R_d is the resistance in the circuit of the electrode. 1, the contact K is closed (injection of holes); 2, holes reach the electrode C (the collector); 3, the contact K is opened; 4, the holes stop reaching the collector.

In strong fields of up to 40,000 V/cm the mobility decreases, as would follow from the theory developed by Landau.[25] At the same time the strong fields increase the concentrations of both holes and electrons.

The effective mass of electrons can be determined from the thermoelectric power α. The effective mass enters into the expression for this quantity

$$\alpha = \frac{k}{e}\left[2 + \ln \frac{2(2\pi m^* kT)^{\frac{3}{2}}}{h^3 n}\right],$$

while the Hall constant R equal to $3\pi/8 \times 1/ne$ does not depend on m^*. Such a treatment leads to values of m^*/m_0 appreciably smaller than unity, of the order 0·3–0·4.

One can also determine m^*/m_0 from the value of the ionization energy ΔE of impurity atoms

$$\Delta E = \frac{13{\cdot}53}{\epsilon^2} \cdot \frac{m^*}{m_0}.$$

For germanium ΔE is about 0·015 eV, while $13{\cdot}53/\epsilon^2$ is about 0·050 eV, so that $m^*/m_0 \doteq 0{\cdot}3$.

Finally, one can estimate the effective mass from the mobility u, starting from the expression

$$\frac{u_1}{u_2} = \left(\frac{m_2^*}{m_1^*}\right)^{\frac{5}{2}} \frac{c_2^2}{c_1^2}$$

where c_1 and c_2 are the interaction constants determining the change in electron energy as the interatomic distance is varied.

For electrons in silicon, as we have seen, $u_1 = 1200$ cm²/V sec, while m^*/m_0 is about 0·6. From this relation we might obtain, assuming that c_1 and c_2 are close to each other in value

$$\frac{(m^*)_{\text{Ge}}}{(m^*)_{\text{Si}}} \doteq 0{\cdot}6\left(\frac{1200}{3600}\right)^{2/5} \doteq 0{\cdot}4.$$

The variation of the energy gap ΔE_0 with temperature depends on the effective mass m^*. For germanium, experimental measurement of the displacement of the absorption edge with temperature leads to $m^*/m_0 = 0{\cdot}35$.

Of all these estimates, which give values of m^*/m_0 of 0·3 to 0·4 the most reliable is the first. However, the value of m^* obtained from it is a function of temperature. If we assume $m^*/m_0 = 0{\cdot}35$, we obtain $\Delta E = 0{\cdot}017$ eV, and the radius of the orbit of impurity electrons $a = 28$ Å. It should be noted, however, that besides these approximate estimates one can obtain an accurate quantitative description of the effective masses for silicon and germanium on the basis of the study of cyclotron or diamagnetic resonance, and magnetic measurements.[26] The value of m^* is a tensor with different longitudinal and perpendicular masses. The equal energy surfaces in the Brillouin zone assume the form of 4 or 8 ellipsoids along the [111] crystallographic axes for germanium, and ellipsoids along the [100] axes for silicon.[27] With regard to holes in the valence band, there are two sets of concentric energy surfaces. Values of the longitudinal and perpendicular effective masses for germanium and silicon have been presented previously in the description of cyclotron resonance.

Determinations of the average time of existence of an excess electron in the free state, τ_0, have shown that in pure germanium this reaches $(0{\cdot}6\text{–}1{\cdot}0) \times 10^{-3}$ sec. The diffusion length L reaches 1 mm and more.

The mean free path, corresponding to the high mobility, is also very large in germanium. It reaches 10^{-5} cm at room temperature.

The presence of impurities in silicon and germanium decreases the temperature dependence of the electrical conductivity and of the concentration of free carriers. For a certain impurity content, not exceeding 1%, silicon and germanium become semi-metals; they preserve a

finite electron or hole concentration even close to the absolute zero. The energy of ionization of impurities in silicon decreases from 0·08 eV to zero in relation to the mean distance between impurity atoms. The experimentally observed relationship for ionization energy ΔE and impurity atom concentration N can be expressed, for silicon, as

$$\Delta E = (0{\cdot}08 - \beta N^{\frac{1}{3}})\,\text{eV}.$$

For $N = 5{\cdot}3 \times 10^{18}$ the ionization energy ΔE reaches zero. An analogous dependence is observed with germanium.

It is possible that the gradual transition to metallic conduction is connected with the large size of the orbit of impurity electrons. The radii, for the orbit with lowest quantum number $n = 1$, are $a_{Si} = 11$ Å and $a_{Ge} = 20$ Å, while for $n = 2$ we obtain $a_{Si} = 44$ Å and $a_{Ge} = 80$ Å. It can be calculated that the orbits with $n = 1$ overlap at concentrations $N = 9 \times 10^{19}$ for silicon and $N = 5{\cdot}7 \times 10^{18}$ for germanium. For orbits corresponding to $n = 2$ the concentrations for which the orbits overlap are 64 times smaller. Under these conditions of strong overlap of their wave functions, the impurity electrons, as has already been indicated, can participate in current flow not only by transferring to the conduction band but by means of direct exchange between impurity atoms.

The preparation of pure germanium is a difficult task. The purity can be judged from the specific resistivity, which should reach several tens of ohm cm if high-quality transistors or solid-state amplifiers of radiofrequency oscillations are to be obtained. In addition to repeated pulling of germanium crystals from the melt in order to obtain particularly pure specimens of germanium and other semiconductors, the method of "zone melting" has also been successfully applied. This method is based on the different distribution of impurities between solid and liquid phases of a material. Since, usually, the concentration of impurities in the liquid phase is higher than in the solid phase, it follows that, on moving a narrow molten zone down a long specimen, the impurities from the edge of the zone that is freezing are gradually "squeezed out" into the moving zone. Eventually, after the process has been repeated many times, the major part of the impurities is concentrated near one end of the ingot. In this way the impurity concentration can be lowered 1000 times or more. The addition of 0·001 to 0·2% of the required impurity can convert such germanium into either *n*-type or *p*-type. A similar kind of treatment can produce a thin surface layer of *p*-type germanium on an *n*-type specimen. Such a specimen shows good rectification.

For transistors and photocells, *n*-type germanium is prepared, on the surface of, or within, which a *p*-type layer is produced. The

boundaries between the layers of opposite carrier sign act as barrier layers. Marked rectifying and photo-effects can be observed at them.

The dependence of current on voltage for a layer situated at the boundary between *n*- and *p*-type germanium has the form shown in Fig. 139. When the reverse voltage exceeds a definite limit, part of the characteristic has a negative slope, the applied voltage falling with increasing current.

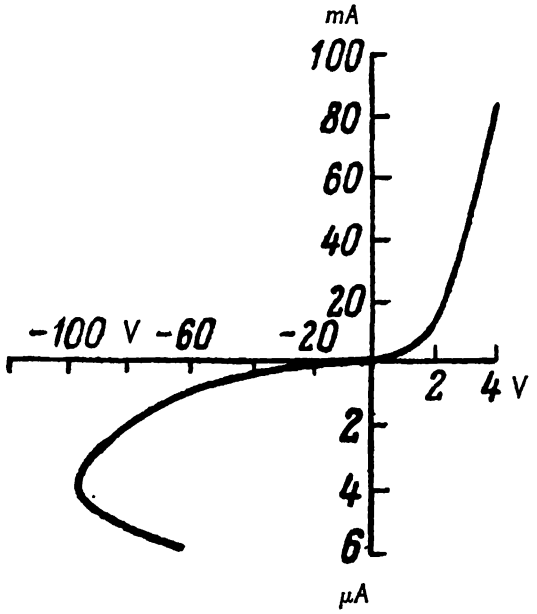


Fig. 139. Voltage–current characteristic of a germanium rectifier.

Germanium devices prepared in this way can be used as detectors, rectifiers, amplifiers, and generators of radiofrequency oscillations. They are not only replacing vacuum electronic devices, but are also opening up new possibilities in high-frequency technology. Germanium detectors, on account of their small capacitance, and of the negligible thickness of the barrier layer (10^{-6} cm) through which the electrons pass with a velocity of 10^{7} cm/sec in 10^{-13} sec, can be used at extremely high frequencies, and particularly in radar apparatus, under conditions in which vacuum devices show poor performance. Their small size and mass, and low cost, are encouraging the wide use of germanium devices in radio technology and radar.

Germanium detectors are also used in computers. Point contact transistors are replacing vacuum triodes in radio broadcasting technology. More powerful devices in which layers of *n*- and *p*-type

germanium are sandwiched together are being used to solve a number of problems in radio-technology and acoustics.

It is clear that new high-power plate rectifiers of alternating current, which have negligible energy losses, have considerable advantages compared with existing rectifiers.

The changes in the electrical properties of germanium under the action of neutrons and fast charged particles—protons, deuterons, and alpha-particles—are of considerable importance.[28]

Electrons with energies greater than $0{\cdot}7 \times 10^6$ eV, and nucleons, displace some of the germanium atoms from their equilibrium positions in the crystal lattice. This gives rise to *p*-type conduction in the germanium. The regularity of the lattice is affected, new scattering centres for electron waves make their appearance, the mobility is reduced, and new sources of free carriers become apparent. Such lattice distortions can be removed and the germanium returned to its original state by long annealing at high temperatures.

Neutron bombardment brings about yet other changes in germanium. They convert germanium atoms into either radioactive isotopes or atoms of neighbouring elements in the periodic system (gallium or arsenic, gallium being formed in amounts three times greater than arsenic). These last effects, clearly, are non-reversible, and the properties of irradiated specimens are permanently altered. For each neutron absorbed by germanium, some 3 electrons disappear in *n*-type specimens and 0·7 holes in *p*-type specimens.

The impurity atoms introduced by irradiation confer *n*- or *p*-type characteristics on germanium depending on their nature. Since in the process of their formation they can be distributed among spaces in the lattice, and not only at lattice sites like chemically introduced impurity atoms, the dissociation energy for them is higher.

Accurately controlled irradiation can be used to compensate for electronic conductivity with a corresponding concentration of acceptors. When this occurs the chemical potential lies approximately at the middle of the energy gap, and the electrical conductivity falls to the intrinsic value for pure germanium. In this case the presence of donors and acceptors can be detected by a reduction in the mobility and a change in its temperature dependence.

The formation of a barrier layer of high specific resistivity at a boundary between *p*- and *n*-type germanium is indicated not only by additional resistance and by rectifying effects, but also by the appearance of photoconduction which is unnoticeable in germanium specimens of high conductivity. Germanium with a junction barrier layer is a sensitive photoconductor for the infra-red region of the spectrum. In this case the photoconduction is accompanied by an amplifying effect.

p–n junction photocells made from germanium have lower efficiencies with solar radiation than silicon devices. This can be understood in terms of the smaller energy gap. Germanium photocells are sensitive to wavelengths close to the absorption edge at 1·6 microns, while the main energy of solar radiation is concentrated around 0·5–0·6 microns. Silicon, for which the absorption edge is at 1·06 microns, approaches more closely the requirements for the use of solar radiation.[29]

We shall deal now with the complex structure of the energy bands in germanium and silicon.

Fig. 140 shows the variation of energy E with wave-number k for the valence and conduction band edges in those cases where the bands are characterized by scalar values of effective mass m^*_- and m^*_+, and where $m^*_- < m^*_+$. Both bounding curves are parabolae.

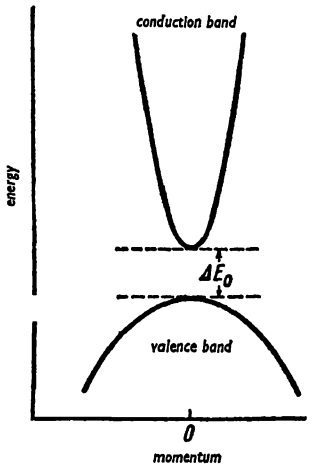


Fig. 140. Simple energy bands.

Fig. 141 shows in the same coordinates the bounding curves for Ge (Fig. 141a) and for Si (Fig. 141b). The left-hand side refers to the crystallographic direction [111], and the right-hand side to the direction [100]. For these cases the valence band is bounded by two curves corresponding to two effective masses for holes. Both these levels coincide at the value $k = 0$, giving rise to a doubly degenerate state. There is also a third level in which the energy is lowered on account of spin–orbit interaction.

The conduction band presents an even more complicated picture in which there are 8 ellipsoids extending along the [111] axes.

There is a maximum in energy at $k = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. The transition from the maximum energy in the valence band to the minimum energy in the conduction band is accompanied by a change in wave vector and so is forbidden for optical transitions. Even so, such forbidden optical transitions are possible if an excited phonon can take part, though they are less probable than transitions satisfying the selection rules.

Fig. 141b refers to Si. Along the [100] directions (on the right) there are found 6 ellipsoidal surfaces for electrons in the conduction band. The [111] direction is indicated on the left.

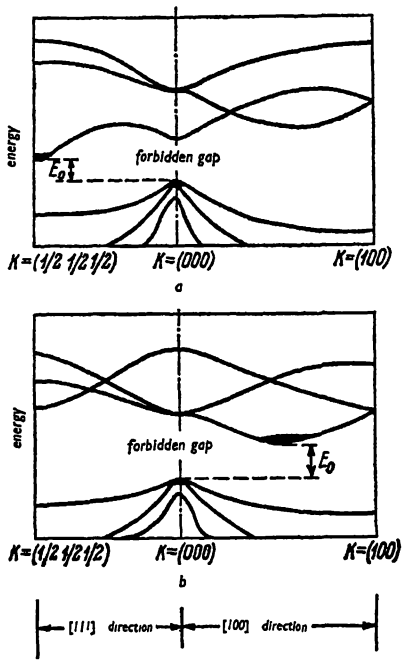


Fig. 141. Energy bands in (a) germanium, and (b) silicon.

For both Si and Ge the maximum in the valence band and the minimum in the conduction band are found at different values of the wave vector k.

Te is also characterized by complex energy surfaces. In this case, as with Ge, there are two types of hole with different effective masses and different mobilities. This is particularly well shown up by the dependence of the resistivity on magnetic field strength, and by the Hall effect.

The complicated energy band structure in Ge and Si gives rise to new kinds of scattering. One now encounters both transitions of electrons from one state to another within a given ellipsoidal region, and

also similar transitions in which the electron transfers from one ellipsoidal region to another. There is reason to suppose that the probability of transitions within a region and between regions is of the same order.

For scattering in which electrons are transferred from one ellipsoidal region to another (this is frequently termed inter-valley transfer) there take part not only acoustic phonons but also optical phonons of higher frequency.

The presence in Ge of two kinds of hole with markedly different mobilities leads to a dependence of the Hall effect on magnetic field strength. If one takes into account both types of hole one obtains good agreement between theory and experiment. Fig. 142 shows the dependence of the Hall coefficient R on the field H for the two temperatures 205° and 297° K, when the field H is perpendicular to the [111] direction.

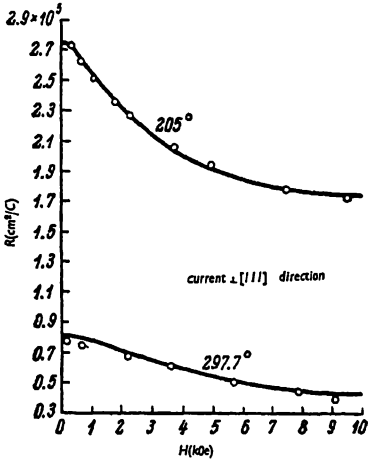


Fig. 142. Dependence of Hall coefficient on magnetic field for Ge. (Continuous curves, calculated values; circles, experimental points).

Similarly good agreement can be obtained for measurements of the change of resistivity ρ in a magnetic field.

For the case of a single scalar effective mass $\Delta\rho/\rho_H$ should be proportional to H^2 and, consequently, $\Delta\rho/\rho_H H^2$ should be constant. In Fig. 143 is shown the variation of $\Delta\rho/\rho_H H^2$ with H for holes of a single type (dashed curve) and for two kinds of hole with a ratio of mobilities of 7·5 and a concentration ratio of 0·02. The experimental results are indicated by circles.

A complex system of energy levels leads to many kinds of transition of electrons and holes from states close to a band edge to other states

and, in consequence also, to a large number of optical transitions, both of the direct type and also of indirect types in which lattice phonons take part.

Careful studies of germanium and silicon have given much information regarding processes of scattering, trapping, and recombination of charge carriers. It appears that in addition to impurities which substitute for atoms of Ge or Si, there are also interstitial impurities which are found within the cells of the lattice, in between its constituent atoms.

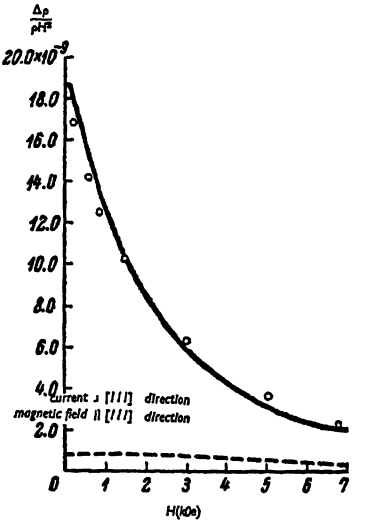


Fig. 143. Dependence of magneto-resistance effect on magnetic field for Ge. (Continuous curve, calculated for two types of hole; broken curve, calculated for one type of hole; circles, experimental points).

We determined previously the energy levels of hydrogen-like impurity atoms; these should lie within a few hundredths of an electron volt of the band edge. An appreciable number of donors and acceptors in fact produce levels which are close to those calculated. However, there are also impurities for which the levels are a few tenths of an eV from the band edge.

Thus, for example, in germanium the impurity Au gives rise either to donor levels at 0·18 eV from the conduction band, or to acceptor levels at 0·05 and 0·15 eV from the edge of the valence band. Atoms of Fe give rise to donors with an ionization energy of 0·27 eV and acceptor levels 0·34 eV from the valence band edge. Such levels, close to the centre of the gap, are also trapping centres (where the probability for capture of minority carriers is much greater than that for majority carriers on account of their electric charge).

If, now, such a deep-lying level has an appreciable probability of bringing together charges of both signs, then it becomes a recombination level, initially trapping one carrier, for example a minority carrier, and then a majority carrier.

TABLE 16

Element	Impurity type[1]	ΔE_D (eV)	ΔE_A (eV)
Germanium			
B	*A*		0·0104
Al	*A*		0·0102
Ga	*A*		0·0108
In	*A*		0·0112
P	*D*	0·0120	
As	*D*	0·0127	
Sb	*D*	0·0096	
Li	*D*	0·0003	
Zn	*A*		0·029
Cu	*A*, *R*		0·25
	A		0·040
	D		0·05
Au	*A*, *R*		0·15
	A, *T*	0·18	
Ni	*A*, *R*		0·23
	A, *T*		
Fe	*A*, *R*		0·34±0·02
	A, *T*	0·27±0·02	
Co	*A*, *R*		0·25±0·01
	A, *T*	0·31±0·01	
Mn	*A*, *R*		0·16±0·01
	A, *T*	0·35±0·01	
Pt	*A*		0·04
Recombination levels	*A*, *R*		~0·25
Frenkel' defects	*A*	0·2	
Silicon			
B	*A*		0·045
Al	*A*		0·057
Ga	*A*		0·075
In	*A*		0·16
P	*D*	0·044	
As	*D*	0·049	
Sb	*D*	0·039	
Li	*D*	0·033	
Au	*D*		0·33
Electron trapping levels	*D*, *T*	0·57	
	D, *T*	0·79	

[1] *A*, acceptor; *D*, donor; *T*, trapping centre for minority carriers; *R*, recombination centre.

In Table 16 are presented data on impurity levels so far known for Ge and Si. It should be borne in mind that at low temperatures neutral

impurity atoms become of importance, as also do lattice distortions caused by plastic deformation (Ge becomes plastic at temperatures above 500° C).

Certain impurities, for example Se and Te, which act as donors, which have not been put to any practical use, are not given in Table 16.

Finally, more accurate data on the relative variation of mobility with temperature for germanium, silicon, and grey tin should be noted:

$$\text{Si} \quad \begin{cases} u_- = 4{\cdot}0 \times 10^9\, T^{-2{\cdot}6} \\ u_+ = 1{\cdot}5 \times 10^8\, T^{-2{\cdot}3} \end{cases}$$

$$\text{Ge} \quad \begin{cases} u_- = 4{\cdot}9 \times 10^7\, T^{-1{\cdot}66} \\ u_+ = 1{\cdot}05 \times 10^9\, T^{-2{\cdot}33} \end{cases}$$

$$\alpha\text{-Sn} \quad u_- = 6{\cdot}3 \times 10^6\, T^{-1{\cdot}5}$$

Tin is the last element of group IV which crystallizes, in one of its modifications, grey tin, in the diamond lattice like silicon and germanium. Grey tin, which is unstable near room temperature, can be formed from white tin at temperatures below 13° C, when it possesses a lower thermodynamic potential than white tin. The transformation can be stimulated by the presence of seeding nuclei of grey tin.

The packing density of the tin atoms in grey tin, with a coordination number 4, is considerably lower than in metallic tin, which crystallizes in a tetragonal lattice with coordination number 6. Accordingly, the transformation from one modification to the other is accompanied by the break-up of the crystals. At low temperatures objects made of tin fall to pieces, and this disintegration can be accelerated by seeding with grey tin. This phenomenon was known even in the Middle Ages and was called "tin pest".

The lattice constant of grey tin is 6·46 Å, and the distance between adjacent atoms is 2·8 Å.

The electrical properties of grey tin have been studied by Blum and Goryunova, and subsequently, in more detail, by Busch, who succeeded in obtaining single crystals.[30, 31]

As would be expected, grey tin is a semiconductor with a still narrower energy gap than germanium (ΔE_0 is about 0·1 eV), and with a high mobility—about 3000 cm²/V sec at room temperature, and 20,000 cm²/V sec at 137° K.

Impurities markedly increase the electrical conductivity of grey tin. Fig. 144 illustrates the dependence of $\log \sigma$ on $1/T$ for pure tin, and for tin with different concentrations of aluminium. The concentration of electrons in pure tin exceeds the limit of degeneracy at temperatures above 70° C.

The calculation of charge concentrations from values for the Hall coefficient is very unreliable on account of the narrowness of the energy gap. For small concentrations of impurities, tin should have mixed conduction even at fairly low temperatures (in its stability region). Accordingly this concentration is more accurately calculated from measurements of the change in the resistivity ρ in a perpendicular magnetic field.

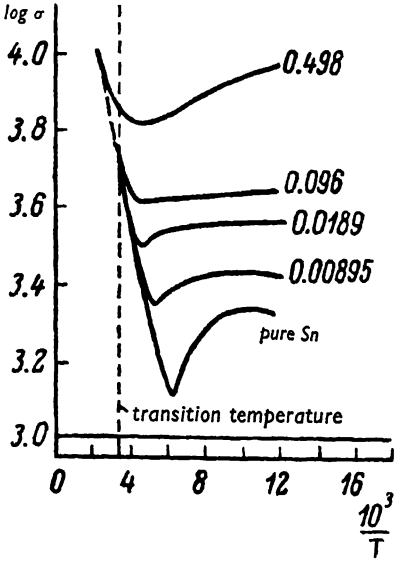


Fig. 144. Temperature dependence of the electrical conductivity of spectroscopically pure α-Sn, and of α-Sn with additions of aluminium (in weight per cent).

For atomic lattices with a single sign of carrier one can state

$$\frac{\Delta\rho}{\rho} = 0{\cdot}48u^2H^2,^{\dagger}$$

while for intrinsic conduction, with equal numbers of holes and electrons,

$$\frac{\Delta\rho}{\rho} = 1{\cdot}75u^2H^2.^{\dagger}$$

Knowing ρ and the mobility u it is a simple matter to calculate the concentration n of free carriers.

Experiments with grey tin show that the mobility dependence $u \propto T^{-\frac{3}{2}}$ is not closely obeyed. This is probably a result of inaccuracies in the

† *Editor's note.* If u is expressed in cm²/V sec and H in oersted, the product uH should be divided by 10^8.

measurements, which are extremely difficult to make, particularly at low temperatures.[32]

Silicon carbide can also be classed with the crystals of Group IV. It is reminiscent in structure of diamond, with the difference that each atom of carbon is bonded to four atoms of silicon, which surround it to form a tetrahedron; similarly, the silicon atoms are surrounded by four atoms of carbon. Silicon carbide occupies a position intermediate between diamond and silicon. Its lattice constant is 4·2 Å, its energy gap 3·5 eV, and its dielectric constant 6·7.[33]

Comparing the properties of the elements of Group IV of the periodic system we note, together with qualitative resemblances, systematic quantitative differences. The properties of these elements are compared in Table 17.

TABLE 17

Material	Atomic weight	a (Å)	ϵ	ΔE_0 (eV)	ΔE (eV)	m^*/m_0	$\Delta E_0\,\epsilon^2$
Diamond	12	3·56	5·7	5·3	0·3	0·7	175
Silicon carbide	(20)	4·2	6·7	3·5	—	—	160
Silicon	28	5·43	12·5	1·12	0·08	0·8	175
Germanium	72·6	5·65	16·0	0·75	0·015	0·25	185
Grey tin	118·7	6·5	50	0·1	—	—	—

The data given in the last column of Table 17 show that it is not only the energy for ionization of impurities that is determined by the dielectric constant of the semiconductor, but also the energy gap, i.e. the ionization energy of the main lattice.

With increase of atomic weight we proceed from diamond, an insulator with an electrical conductivity 10^{-14} ohm^{-1} cm^{-1}, to silicon with σ equal to 5×10^{-6} ohm^{-1} cm^{-1}, to germanium with σ equal to 10^{-2} ohm^{-1} cm^{-1} and to grey tin with a conductivity of 1000 ohm^{-1} cm^{-1}. The diamond-like lattice of grey tin, with four valence bonds to each atom, is already unstable at room temperature and transforms to a more closely packed lattice in which each tin atom is surrounded by six neighbouring atoms, and is bonded to them by metallic bonds. The next element in Group IV, lead, is a typical metal.

The purely covalent bond of diamond becomes weakened on going to lower elements in the group, the lattice becomes looser, and finally, in tin and lead, the bonds become purely metallic.

In addition to the crystal lattices considered, compounds between elements of the third and fifth columns of the periodic system also crystallize in diamond-like lattices with valence bonds. Examples are AlSb, InSb, GaSb, AlP, AlAs, etc.[34, 35]

All these compounds are analogous in their physical properties to the group: diamond, silicon, germanium. The mobility of electrons at 20° C reaches 80,000 cm^2/V sec with InSb. The lattice constant of InSb is so close to that of grey tin that it can act as a seeding catalyst for the crystallization of grey tin, and for the transformation of white tin into grey tin at temperatures below 13° C. Certain compounds between the second and sixth columns, for example HgSe, have similar properties; the mobility of electrons in HgSe is greater than 10,000 cm^2/V sec.

Data regarding compounds between the third and fifth groups, with diamond-type lattices, are compared in Table 18. With some of them it is possible to obtain *p–n* junctions, and to use them in place of germanium. Although, judging from their energy gap, the first three compounds fulfil the requirement of a close match between their region of photoconductivity and the region of maximum energy in the solar spectrum better than silicon, it has not proved possible to prepare photocells of high efficiency from any of them.

TABLE 18

Material	Melting point (°C)	Mobility		Diffusion length (microns)	Energy gap ΔE_0 (eV)
		Electrons cm^2/V sec	Holes cm^2/V sec		
AlSb	1080	—	200	20	1·60
GaAs	1240	1,200	—	—	1·35
InP	1070	3,400	650	400	1·25
GaSb	720	5,000	650	—	0·67
InAs	936	23,000	100	—	0·33
InSb	523	80,000	700	—	0·18

There are also more complex compounds with crystal lattices of the diamond type, among them minerals, for example, chalcopyrite and stannite. The first of these, with the chemical formula $CuFeS_2$, gives rise to a large series of analogues in which univalent copper can be replaced by Ag or Au, tervalent iron by Al, Ga, In or Tl, and sulphur by Se or Te. The energy gap in these compounds has values from 2·5 eV for $CuAlS_2$ to 0·1 eV for $AgTlTe_2$.[36, 37]

Stannite, Cu_2FeSnS_4, can also be considered as the source of a series of compounds in which, however, iron is replaced by divalent metals and not by tervalent ones, as in chalcopyrite.

In some of these more complex compounds it is possible to produce *p–n* junctions. Mobilities of some hundreds of cm^2/V sec have been measured for such compounds, and it is probable that with better

purification much higher values will be obtained. Zhuze has found mobilities up to 5000 cm^2/V sec in some compounds of the stannite type.

A further example of a crystal with purely valence bonds is boron. Its energy gap is 1·1 eV, the energy of ionization of impurities is about 0·5 eV, and its dielectric constant 6·2. It can be concluded from this that $m^*/m_0 = 0{\cdot}7$. The sign of the carriers in boron is positive.

Molecular lattices

The elements of Group VI form molecular lattices. The electrical properties of solid and liquid oxygen have not been studied. It is only known that oxygen is an insulator.

Sulphur, one of the best insulators, shows marked photoconduction with a maximum in the blue part of the spectrum. The electrical conductivity of sulphur in the dark is of the order of 10^{-18} ohm^{-1} cm^{-1}. When illuminated it increases by a factor of thousands or even millions.

Sulphur crystallizes in a number of modifications (rhombic and monoclinic). Crystalline rhombic sulphur melts at 113° C. If liquid sulphur heated to a high temperature is rapidly quenched, one can obtain amorphous sulphur at room temperature. This subsequently crystallizes slowly.

Despite its small atomic weight, the thermal conductivity is very small, not only for amorphous sulphur, but also for the crystalline form ($1{\cdot}1–1{\cdot}4 \times 10^{-3}$ cal/cm sec deg) depending on the crystal direction.

It is possible that the very small value of thermal conductivity can be explained in that sulphur crystallizes in complex molecular lattices built up of eight-sided rings (128 atoms in the cell). Accordingly, the fundamental thermal vibrations of the lattice are vibrations of the rings relative to each other, and are of low frequency. The strongly bonded atoms within the rings, on the other hand, produce vibrations of so high a frequency that they would only be in evidence at high temperatures, when sulphur is no longer in the solid form.

The most thoroughly studied of the succeeding elements in the sixth column is selenium.[38–41]

Selenium, like sulphur, has more than one modification: red, insulating selenium, and hexagonal, grey, semiconducting selenium. Both modifications show marked photoconduction. The grey modification, which melts at 220° C, is used in rectifiers and photocells.

When molten selenium solidifies, an amorphous mass is first obtained which, after long annealing above 170–180° C, transforms to the crystalline state. While amorphous selenium has a resistivity of the order of

10^{13} ohm cm, the resistivity of grey crystalline selenium is 10^5 ohm cm and is not greatly altered on adding impurities.

The thermoelectric power of ordinary selenium at room temperature is 600 microvolts/deg; it increases with temperature.

Selenium is a typical *p*-type semiconductor. Impurity ions as well as holes can migrate in the selenium lattice, in particular the tervalent ion of thallium. At temperatures near the melting point of selenium the mobility of thallium ions reaches $1{\cdot}7 \times 10^{-6}$ cm^2/V sec. For a long time attempts at detecting and measuring a Hall effect were unsuccessful, and it is only by the application of the most sensitive methods that the hole mobility has been determined as 0·1–1 cm^2/V sec. The concentration of holes is about 10^{15}–10^{16} cm^{-3}.

The energy gap, as determined from the absorption edge, is 0·8 eV. Hydrostatic pressure decreases both the energy gap and the separation between impurity levels and the band edge. The resistivity of selenium decreases by 30% on increasing the pressure to 1000 atmospheres.

From the temperature variation of the electrical conductivity of selenium it can be concluded that the acceptor levels are 0·1 eV from the upper edge of the valence band. However, the weak dependence of the electrical conductivity of selenium on impurity content, compared with that for other semiconductors, compels one to assume that, in addition to foreign atoms, dissociated atoms of selenium itself might act as acceptors. These would be present in but small numbers since the large majority of selenium atoms are built into the long chains which form the hexagonal crystals of grey selenium.

Over many decades hundreds of papers devoted to selenium have been unsuccessful not only in developing a clear picture of the processes occurring in this material, but even in obtaining agreement in describing its properties. For example, the dependence of photoconductivity σ_p on light intensity L has been given by different authors as $\sigma_p \propto L^{\frac{1}{2}}, L, L^{\frac{3}{2}}, L^2$, while the temperature coefficient of the electrical conductivity has been found to be both negative and positive with ionization energies of from 0 to 0·2 eV.

It can be assumed that all these discrepancies can be explained in terms of the special structure of selenium as a polymer.

Molten selenium consists partly of closed rings of eight atoms, as with sulphur, and partly of long chains of hundreds or even thousands of atoms. Crystallization requires a rearrangement of this structure, and the formation of spiral chains, arranged parallel to each other, in which the atoms occupy regular crystallographic sites. It is clear that this process takes place slowly. During its course the density of the crystallizing mass increases. Crystallization can be accelerated by the presence of catalysts, thallium, alkali metals and amines, while iodine,

on the other hand, slows it down. It can also be assumed that the polymeric crystal of selenium as formed contains within it a large number both of distortions and of individual atoms which have not been incorporated into the chains of selenium atoms. The regrowth of the selenium chains with the formation of small ordered crystallites creates large mechanical strains in them, while a semi-amorphous layer with enhanced electrical resistivity is left at the boundaries. As with other polymers, any processes which are accompanied by structural change will occur slowly in selenium, and will lead to hysteresis effects.

The presence of impurity ions in selenium with their comparatively large mobilities leads to structural changes on passing an electric current or on producing additional ions by the action of light or strong fields, and these changes are annulled slowly. The thermal conductivity κ of amorphous selenium is about 4×10^{-4} cal/cm sec deg, while that of finely crystalline selenium is $2{\cdot}6 \times 10^{-3}$ cal/cm sec deg. The small value of κ, in spite of a low atomic weight, can be explained, as in the case of sulphur, in terms of the molecular structure of selenium.

Selenium rectifiers are very widely used. An amorphous layer of selenium, 60–80 microns thick, is formed on a substrate of iron or nickel-plated iron. No barrier layer is formed at this interface. An alloy, containing cadmium, tin, and bismuth, serves as the other, rectifying, electrode. This alloy has a smaller contact potential than selenium, and forms a barrier layer of enhanced resistivity on the selenium which has been transformed to the grey hexagonal modification by heat treatment. However, this layer shows only poor rectifying properties.

In order to prepare technically acceptable rectifiers with rectification factors of the order of some hundreds or thousands (see Fig. 145), it is necessary to carry out a forming process. An appreciable potential is applied for a definite time across the barrier layer at a high temperature. After such a forming treatment, X-ray analysis indicates the presence of a layer of *n*-type cadmium selenide at the boundary with the *p*-type selenium. It can be assumed that the good rectifying properties of selenium rectifiers are determined by the junction between the *n*-type layer of CdSe so produced and the bulk of the *p*-type selenium.

Although selenium has a high resistivity, the energy losses in rectifiers due to Joule heating are relatively low on account of the small thickness of the layer. The efficiency of good rectifiers reaches 80%. With the passage of time there is a deterioration in the properties of selenium rectifiers. A further forming treatment can to some extent restore them.

Selenium photocells have been widely used in the past, but are now being replaced by other, more sensitive devices made from thallium or

cadmium sulphide or, for use in the infra-red, from lead sulphide, selenide, or telluride.

Selenium begins to absorb light strongly in the infra-red region of the spectrum at a wavelength of 1·6 microns. When supported on rock salt or quartz, layers of selenium are used as filters transmitting infra-red radiation.

In the amorphous state, and also in the liquid state which is its direct continuation above 220° C, selenium shows *p*-type conduction, as does the crystalline grey modification.

The energy gap decreases with increasing temperature by 0·14% per degree. At room temperature it is 2·2 eV for amorphous selenium, as opposed to 0·8 eV for grey crystalline selenium. The coefficient of volume expansion of amorphous selenium is $1{\cdot}5 \times 10^{-4}$ per degree.

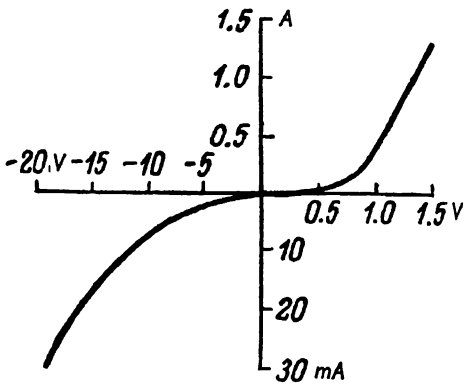


Fig. 145. Voltage–current characteristics of a selenium rectifier with an active area of 15 cm².

Barrier layer photocells made from selenium are still of some importance. Their spectral sensitivity is close to that of the human eye while their frequency response is fairly constant up to some thousands of cycles per second. The electromotive force produced by bright sunlight can reach 1 volt. The dark resistance of the barrier layer is usually 10,000–20,000 ohms. On illumination it falls to 1000–500 ohms, and slowly returns to its original value in darkness. The efficiency η (the ratio of the electrical energy obtained on illumination to the radiant energy falling on the photocell) does not exceed 0·05% under such conditions. Kolomiets and Maslakovets[42] have prepared photocells from thallium sulphide for which η reaches 1%, while for silicon *p–n* junction photocells it reaches 11%.

The next element in Group VI is tellurium, which in many respects is analogous to selenium. It crystallizes from the amorphous phase even at room temperature, and can readily be obtained in single crystal form. On crystallization the correct structure is established slowly, and, accordingly, the density of crystalline tellurium shows a slow increase with time.

In contrast to the idea held previously that tellurium was metallic, it is now considered a semiconductor with a comparatively narrow energy gap, 0·34 eV, and usually with *p*-type conduction. The mobility of electrons in tellurium at room temperature reaches 1500 cm^2/V sec, and the hole mobility 1200 cm^2/V sec. With increasing temperature the mobility of holes decreases more slowly than that of electrons, and above 240° C it exceeds the electron mobility. Measurements on tellurium of the increase of resistivity in a magnetic field lead to the conclusion that two kinds of hole with different masses are present.

Hydrostatic pressure decreases the energy gap to 0·2 eV at 11,000 atmospheres, and to 0·1 eV at 17,000 atmospheres. At a pressure of 30,000 atmospheres the energy gap decreases to zero and tellurium becomes a metal. At 45,000 atmospheres tellurium transforms into a new metallic modification.

The elements of Group V, phosphorus, arsenic and one of the modifications of antimony, are semiconductors, and in many respects duplicate the properties detailed above for elements of Group VI.

Phosphorus in its white modification has a tetrahedral structure. Above 200° C it gradually transforms into the red amorphous modification with a structural network in which the angles are similar to the tetrahedra of crystalline phosphorus. At temperatures above 500° C, red phosphorus, over the space of a few days, gradually rearranges itself into a system of micelles, which at first are buried in the amorphous medium. Under a pressure of 12,000 atmospheres phosphorus forms a semiconducting black modification which shows rectifying properties.

Arsenic, at low temperatures, has a red crystal structure, analogous to white phosphorus. Above 200° C a black amorphous modification is formed which at 300–500° C begins to change into crystallites of the metallic modification.

Antimony, when sublimed *in vacuo* on to a cold substrate, forms an amorphous layer which gradually crystallizes at room temperature. In the presence of chlorine, the antimony slowly transforms into non-conducting explosive antimony.

The electrical properties of the elements of Group V have not been studied systematically, and have not as yet found any technical application.

Sulphides, selenides and tellurides

The sulphides, selenides and tellurides of metals are a group of semiconductors with varied properties, ranging from the insulator, pure cadmium sulphide, to lead and bismuth tellurides which, in conductivity, are close to metals. These last compounds are distinguished by high carrier mobilities (1000 cm^2/V sec and more).

Cadmium sulphide, which is readily obtained in the form of single crystals has an electrical conductivity of the order of 10^{-10} ohm^{-1} cm^{-1}. This is markedly increased not only by illumination but also by the passage of individual alpha-particles and other nucleons. Accordingly, cadmium sulphide is used as a counter for nuclear particles and also as a photoconductor, its sensitivity for visible light exceeding that of all other materials.

The formation of a p–n junction in cadmium sulphide leads to a photocell which generates electrical energy when illuminated. Although pure crystals of CdS have a conductivity of the order of 10^{-10} ohm^{-1} cm^{-1}, excess of cadmium, and of other impurities also, raises σ to 1000 ohm^{-1} cm^{-1}. The dielectric constant ϵ is 11·6, and the energy gap ΔE_0 is 2·1 eV. The impurity activation energy ΔE for low concentrations is 0·04 eV; this decreases to zero as the concentration is increased to 10^{18} cm^{-3}. The electron mobility at room temperature is 210 cm^2/V sec and this increases by a factor of 10 on lowering the temperature to 40° K. The thermoelectric power α decreases with increasing concentration n of free carriers according to the law $\alpha = A - B \ln n$. The effective mass of electrons m^* is $2m_0$; the current carriers are usually electrons.

Thallium sulphide is also one of the most sensitive photoconductors in use. Photocells with a sensitivity of 10,000 microamps/lumen can be prepared from it. Thallium sulphide photocells, however, only maintain their sensitivity for a few months. More stable though rather less sensitive photocells have been made from silver sulphide by the Physical Institute of the Academy of Sciences of the Ukrainian SSR.[43]

Cuprous sulphide of high p-type conductivity is used in rectifiers, with aluminium or magnesium as the rectifying electrode. These rectifiers, which were developed by Kurchatov and Dunaev,[44] require, like selenium rectifiers, a preliminary forming treatment. They have advantages in respect of allowed current densities, 5–10 A/cm^2, instead of 0·05–0·1 A/cm^2 obtained with selenium and cuprous oxide rectifiers. The defects of the sulphide rectifier are its low efficiency, which does not exceed 65–70%, and difficulties in using them connected in series (this is due to an appreciable lag in the fall of reverse current when the applied potential changes sign). In a circuit with several series-connected

rectifiers, after a change in current, one element will change its resistance more rapidly than the others, so that the potential drops become different and for some rectifiers will exceed the permitted limit, resulting in breakdown.

Lead sulphide, selenide and telluride furnish layers which are photosensitive in the infra-red region of the spectrum. The limit of photosensitivity for lead sulphide is at 3 microns at room temperature, 4 microns at liquid air temperature, while at the temperature of liquid hydrogen it is at 4·5 microns. For lead telluride the corresponding limits of photosensitivity are at 4·5, 6 and 7 microns. The absorption of oxygen by layers of these compounds has a great influence on the magnitude and spectral distribution of their photoconductivity.[45]

Sensitivity in the infra-red region of the spectrum is also shown by cuprous telluride (to 2·2 microns), and by the tellurides of silver (to 3 microns), zinc (to 2·6 microns), cadmium and mercury (to 3·2 microns), antimony (to 2·6 microns) and bismuth (to 4 microns). The radiation-enhanced conductivity returns to its normal value in darkness within a few ten-thousandths of a second. If the temperature is raised the time taken for recovery is reduced to a few micro-seconds.

Photosensitivity in the far infra-red region of the spectrum shows a certain lack of conformity with the electrical characteristics of the above compounds. Their energy gaps, as indicated by the temperature variation of their dark conductivities, can be estimated as about 0·5 to 1 eV. Consequently, a photoeffect at a wavelength of 5 microns, i.e. at $h\nu = 0{\cdot}25$ eV, cannot be ascribed to the excitation of electrons from the valence to the conduction band. Furthermore, the high photosensitivity is not compatible with a mechanism by which electrons are excited from impurity levels, the number of which does not exceed even 1%, while the energy to excite electrons from these levels is too small, of the order of 0·1 eV and not 0·25 eV.

It is possible to divide the absorption spectra of PbS, PbSe, and PbTe into two parts: a sharp fall-off in absorption at shorter wavelengths; and, in the far infra-red part of the spectrum, a slowly decreasing tail showing a maximum (see Fig. 146). If one ascribes the first part to the excitation of electrons from the valence to the conduction band or to exciton states, one can estimate the energy gap ΔE_0 by extrapolating the sharp fall-off in absorption. This gives the values

	ΔE_0 (eV)
PbS . .	1·04
PbSe . .	0·83
PbTe . .	0·62

The last result exactly coincides with that obtained from the temperature variation of Hall coefficient for PbTe.

So far as photosensitivity is concerned, one must take into account the decisive influence of absorbed oxygen. This can increase the photosensitivity a hundredfold, while the infra-red limit is shifted some 1–2 microns toward longer wavelengths.

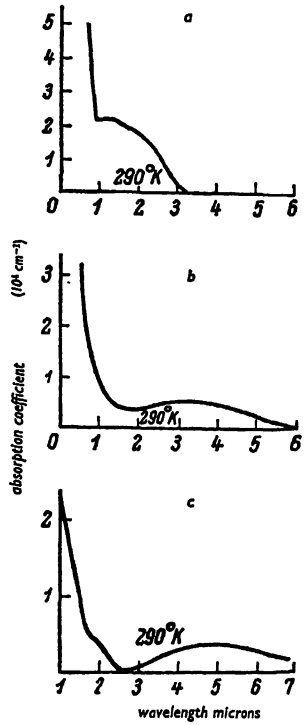


Fig. 146. Absorption spectra at room temperature (according to Gibson): *a*, PbS; *b*, PbSe; *c*, PbTe.

One of the explanations advanced assumes that oxygen in the form of negative ions is absorbed into intercrystalline layers on the surface of the crystallites. The negative charge of the ions produces a barrier layer in the *n*-type crystal which raises the edge of the conduction band and increases the resistance of the layer. Light, by freeing electrons from the oxygen ions, then lowers the resistance of the barrier layers and increases the electrical conductivity of the specimen. An argument against this explanation is that the most sensitive layers are *p*-type and not *n*-type.

Another explanation requires the existence of barrier layers of high sensitivity between n- and p-type crystallites.

Whatever the nature of the high infra-red sensitivity may be, it cannot be ascribed to excitation of electrons across the energy gap. So far the role of excitons in these materials has not been studied.

Besides the photoelectric applications, there have been attempts to make use of lead sulphide, selenide and telluride as radio detectors and also as rectifiers, to replace germanium and silicon. Excess of metal over the stoichiometric composition leads to n-type conduction in these compounds, while excess of sulphur, selenium or tellurium gives p-type material. Accordingly one can obtain with these materials a junction between p- and n-type regions which acts as a rectifier.

Finally, the large mobilities of electrons and holes, together with the small thermal conductivity of the crystal lattice, leads to the possibility of employing these compounds in thermoelectric apparatus. At room temperature the electron mobility u_- in PbS is 400–500 cm^2/V sec, while with PbTe it reaches 2000 cm^2/V sec. The hole mobility u_+ for PbTe at room temperature is appreciably less than u_-.

The mobility of electrons falls with increasing temperature much more rapidly than would follow from the theory, both for an ionic lattice (for which at temperatures above the Debye temperature, $u \propto T^{-\frac{1}{2}}$) and for a covalent lattice (for which $u \propto T^{-\frac{3}{2}}$). It has been shown by experiment that the variation with temperature is much larger than either of these.[46] Thus with PbS and PbTe, for example, the mobility varies with temperature as T^{-3}. The mobility variation with PbTe becomes reduced at temperatures below 150° K and then approximates to $u \propto T^{-\frac{3}{2}}$. In this case scattering of electrons by impurities and lattice distortions, the number of which does not vary with temperature, begins to predominate over scattering due to thermal vibrations of the lattice.

The ionization energy of excess atoms of Pb or S in PbS is zero, since the free carrier concentration remains almost unchanged right down to 2° K. Lead sulphide in this range should be considered as a semi-metal. However, on raising the temperature, electrons and holes are formed by excitation across the forbidden gap of the semiconductor. The larger the concentration of excess atoms of lead or sulphur the higher the temperature at which the intrinsic conductivity of the crystal lattice becomes apparent (Fig. 147).

The thermal conductivity is of great importance for thermoelectric applications. The lattice thermal conductivities for PbS, PbSe and PbTe vary from 7×10^{-3} cal/cm deg sec for the first to 5×10^{-3} for the last. The thermal conductivity of Bi_2Te_3, which is also used in thermoelements is smaller still, about $2{\cdot}5 \times 10^{-3}$ cal/cm deg sec.

The conduction electrons, however, introduce an additional thermal conductivity. In accordance with the Wiedemann–Franz law, and as has been confirmed by experiment, the electronic contribution to the thermal conductivity is, for non-degenerate electrons at room temperature, about 0·001 cal/cm deg sec for each 1000 $ohm^{-1}\,cm^{-1}$ of electrical conductivity. In the degenerate region the value is about 0·0016 cal/cm deg sec for each 1000 $ohm^{-1}\,cm^{-1}$. As was shown in the section on thermal properties, the effect of the carriers on the thermal conductivity becomes more important with the onset of intrinsic conduction.

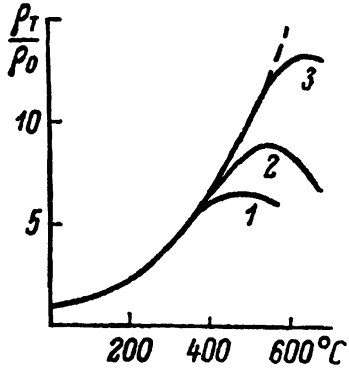


Fig. 147. Temperature dependence of the resistivity for lead sulphide. 1, $n = 5{\cdot}9\times10^{18}$ cm^{-3}; 2, $n = 7\times10^{18}$ cm^{-3}; 3, $n = 1{\cdot}9\times10^{19}$ cm^{-3}.

The thermoelectric power α for PbTe can be well described by either a theoretical formula corresponding to the experimentally determined relationship $u \propto T^{-\frac{1}{2}}$ with $m^* = m_0$, or by the formula for covalent lattices, with $m^* = 0{\cdot}37m_0$

$$\alpha = \frac{k}{e}\left[1 + \ln\frac{2(2\pi m_0 kT)^{\frac{3}{2}}}{h^3 n}\right],$$

or

$$\alpha = \frac{k}{e}\left[2 + \ln\frac{2(2\pi\times 0{\cdot}37 m_0 kT)^{\frac{3}{2}}}{h^3 n}\right].$$

With PbS, for which the impurity levels coincide with the edges of the forbidden band, and for which the chemical bonds are markedly ionic, α is somewhat smaller than the theoretical value.

The compounds PbS, PbSe and PbTe form solid solutions with each other; so also do the group of compounds Sb_2Te_3 and Bi_2Te_3 and analogous compounds of the metals Sb and Bi with sulphur and selenium. The second group have markedly anisotropic lattices of a layer type, and have a high electrical conductivity with a reduced thermal conductivity. They are, accordingly, good materials for thermoelements.

The compound In_2Te_3 has a ZnS type lattice similar to that of diamond, but with a predominantly ionic bond. One third of the sites in the lattice which should be occupied by In are vacant. As a result the crystal has a loose structure and an extremely large number of scattering centres, which lower the mobility below the limit of detection. In agreement with this low mobility the electrical conductivity of In_2Te_3 is small. The energy gap is in the range of 2–4 eV.

The energy gap of the compound CdTe is 1·5 eV.

Ag_2Te has two modifications; the high-temperature α- and low-temperature β-forms both have high electrical conductivities. The α-modification shows predominantly electrolytic conduction; the β-modification is noteworthy for an extremely high electron mobility of the order of 4000 cm^2/V sec, which indicates the covalent nature of the bonds in the β-Ag_2Te lattice.

Oxides

Both p-type (Cu_2O, UO_2) and n-type (ZnO, TiO_2) materials are to be found among the oxide semiconductors.

Cuprous oxide, from which the first rectifiers and photocells were prepared, has been studied in a particularly detailed fashion. In experiments which have already been mentioned, Zhuze and Kurchatov established that the energy gap for Cu_2O is about 1·5 eV, and that the impurity levels, which are acceptors, are at a distance of 0·3–0·6 eV from the edge of the valence band. These conclusions are confirmed by optical data: the extrapolated limit of infra-red absorption and also the commencement of the internal photoeffect are close to 0·8 microns, which corresponds to a gap width 1·55 eV. In addition there is a second region in which a photoeffect is observed at wavelengths of the order of 2 microns (0·63 eV).

At room temperature and above, up to 300° C, cuprous oxide is usually in a non-equilibrium state. Only at temperatures of 800–1000°C does it rapidly come to equilibrium with oxygen in the surrounding medium. Absorption of oxygen increases the number of acceptors N, and enhances the p-type conductivity. The number of acceptors N increases with the partial pressure of oxygen

$$N \propto p^{\frac{1}{4}}.$$

It does not follow from this, however, that atoms of oxygen passing into cuprous oxide act as acceptors. It is probable that the oxygen atoms absorbed in excess convert some of the monovalent copper ions into divalent ions. Thermal motion subsequently carries electrons to these ions from the valence band. Some such mechanism of oxygen absorption lies behind the above-mentioned connection between N and p.

It is also possible that ions of copper diffusing to the surface combine there with oxygen, leaving empty sites in the lattice.

The dependence of the electrical conductivity of cuprous oxide on oxygen pressure has raised doubts regarding the validity of the conclusions of Zhuze and Kurchatov concerning the possibility of determining the temperature variation of electrical conductivity.

If at high temperatures we are dealing not with a metastable state, but with one in thermodynamical equilibrium, then the concentration of impurity centres can vary not only with oxygen pressure but also with temperature, according to the equation

$$N = Ap^{\frac{1}{2}} \exp\left(-\frac{E_k}{kT}\right)$$

where E_k is the energy expended on absorbing one atom of oxygen. If this excitation energy for absorbing oxygen is available, then the electrical conductivity σ, in the range of temperature for which equilibrium is established with the surrounding medium during the time of measurement, is determined by the expression

$$\sigma = CN_{\text{eff}}^{\frac{1}{2}} \exp\left(-\frac{\Delta E_0}{2kT}\right) = CA^{\frac{1}{2}} p^{\frac{1}{4}} \exp\left(-\frac{\Delta E_0 + 2E_k}{2kT}\right).$$

Accordingly the slope of the plot of $\ln \sigma$ against $1/T$ in the experiments of Zhuze and Kurchatov should have given, one would expect, not the energy gap ΔE_0 but a larger quantity $(\Delta E_0 + 2E_k)$. This conclusion, however, is not supported by the optical properties of cuprous oxide. The absorption edge and the photoeffect limit lie at still higher values of $h\nu$ than that corresponding to the value obtained by Zhuze and Kurchatov of $\Delta E_0 = 1{\cdot}44$ eV. Later, more careful measurements by the same method gave $\Delta E_0 = 1{\cdot}56$ eV. Apparently the energy E_k required for the absorption of oxygen is very small in comparison with $\Delta E_0/2$.

The mobility of holes in cuprous oxide at room temperature is 60–80 $\text{cm}^2/\text{V sec}$. The mobility increases at low temperatures and decreases at high temperatures more rapidly than the theoretical variations $u \propto T^{-\frac{1}{2}}$ for covalent, or $u \propto T^{-\frac{3}{2}}$ for ionic, semiconductors. The experimental variation lies between $u \propto T^{-\frac{1}{2}}$ and $u \propto T^{-\frac{3}{2}}$. At high temperatures the negative exponent is higher still.

As has been shown by Pekar, the variation of mobility with temperature is in good agreement with his theory of polarons, according to which the moving holes are surrounded by a polarized region of the crystal. Absolute values of the mobility, however, do not agree with the theory.

Current flow in cuprous oxide obeys Ohm's law up to fields of the order of 10^4 V/cm. At higher fields there commences an exponential increase in conductivity which is connected with an increase in the concentration of carriers, their mobility remaining unaltered.[47]

The dielectric constant of cuprous oxide is 8·75.

The photoconductivity of cuprous oxide has been studied by a large number of workers. The maximum photosensitivity is displaced towards short wavelengths with reduction in temperature.

A. V. Ioffe and I have shown that over a wide range of wavelengths around the maximum the photoconductivity is directly proportional to the energy absorbed (Fig. 148), and that the effect of light and of strong fields leads to an increase in the concentration of holes and not in their mobility. At low temperatures there is considerable inertia in the restoration of the dark conductivity.[48]

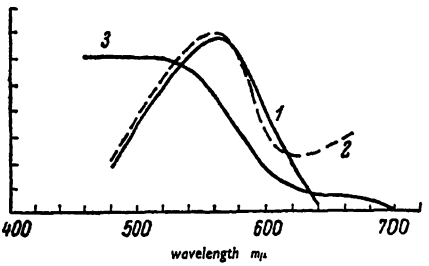


Fig. 148. The relationship between photoconductivity (1), and absorbed energy (2) for cuprous oxide. The dependence of the absorption coefficient on wavelength (3).

The photocurrents which are observed along the direction of illumination give rise to photo e.m.f.s which can reach 0·2 volt. Kikoin and Noskov obtained photo e.m.f.s of up to 15 volt at liquid air temperature in a magnetic field and gave a completely satisfactory theory of this phenomenon. The Kikoin–Noskov photomagnetic effect is now much used for the study of systems of energy levels in semiconductors.[49]

A. V. Ioffe and I have observed in the strongly absorbed blue region of the spectrum not only an increase, but also a decrease in electrical conductivity (negative photoeffect). This can be explained if one takes into account the diffusion of electrons and holes. The theory of such bipolar diffusion and of the photo e.m.f.s observed in cuprous oxide has been developed and confirmed by the careful experiments of Lashkarev.[50] He has also studied the infra-red radiation which arises in cuprous oxide during recombination.

The mechanism of the photoeffect in cuprous oxide has been studied in great detail and explained by Zhuze and Ryvkin.[51, 52] By observing

the rate of growth of photoconductivity on illumination, and its decay on cutting off the light, they were able to determine the lifetime of holes, and to show that the photoeffect in cuprous oxide is a two-step process. The first step is the formation of excitons, as theoretically predicted by Frenkel'. The excited electron is still attached to a hole and can move without increasing the electrical conductivity. Only later, when the exciton, diffusing through the crystal, encounters an acceptor level occupied by an electron, can it give up its energy, and raise the electron into the conduction band, from which it soon falls into a trapping level. Thermal motion brings an electron from the valence band to the freed acceptor level leading to an increase in the hole conductivity.

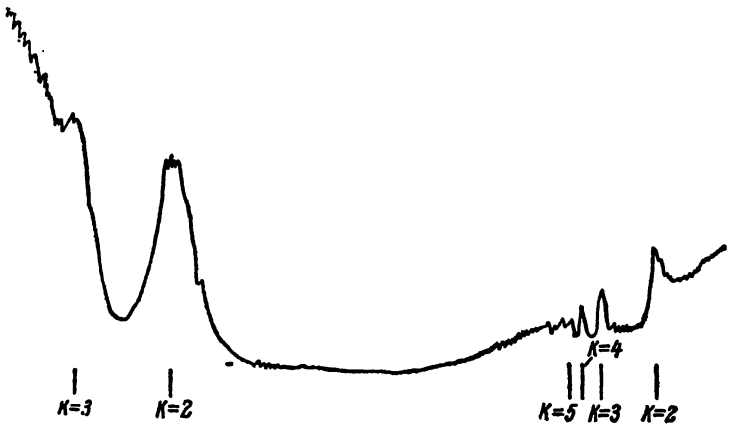


Fig. 149. Microphotograph of the main absorption edge for cuprous oxide (according to Gross).

The conclusions of Zhuze and Ryvkin were confirmed by Gross and Karriev.[53] Employing an apparatus of extremely high resolving power, these workers observed in the absorption spectrum of cuprous oxide, close to the limit of strong absorption, a series of lines, which are reminiscent of the hydrogen spectral series (Fig. 149). Gross has observed analogous absorption spectra with certain other materials also. The relationship between the frequencies of these lines agrees with the frequency relationship of the Balmer series for hydrogen. We are apparently dealing here with discrete excitation levels which, on increasing the frequency of the radiation, go over into the ionization level of a singly charged ion situated in a medium of high dielectric constant. The data of Gross are in particularly good agreement with Pekar's theory of polarons and reproduce a polaron spectrum. Detailed examination of the absorption spectrum of cuprous oxide, shown in Fig. 149,

has revealed, in addition to two series of lines, a reduction of intensity in the continuous spectrum beyond the limit of the first (left-hand) series. This reduction indicates, apparently, the existence of two conduction (or valence) bands which are either close together or partially overlapping.

The second hydrogen-like series in cuprous oxide, according to the views of Gross, is, like the first series, due to excitons. The limit of the second series is displaced relative to the limit of the first series by 0·131 eV.[54]

Rectifiers made from cuprous oxide have been known for over 25 years. They are prepared from copper plates, which are coated with a layer of oxide at 1050° C. The oxide is subsequently saturated with oxygen at a temperature of about 600° C. The black layer of cupric oxide formed on the surface is removed by dissolving it in acid and a metallic electrode applied. This treatment produces a barrier layer at the boundary between the copper plate and the layer of oxide grown on it. No such barrier layer is formed at the interface with the applied, second, electrode. The difference in contact potential between copper and cuprous oxide is insufficient to produce a large enough curvature of the bands for the conduction band edge to approach the Fermi level, and for the oxide to be converted from *p*-type to *n*-type. Accordingly, it is necessary to assume that the layer of oxide in immediate contact with the copper, having an excess of copper and other impurity atoms, is an *n*-type semiconductor. The conditions necessary for rectification then exist at the boundary between this *n*-type cuprous oxide and the *p*-type material.

It should be noted that a simple contact of cuprous oxide with any metal possessing a smaller contact potential does not give rise to any appreciable rectification, even though a layer of high resistance is produced. However, a contact with an *n*-type semiconductor (titanium dioxide or zinc oxide) leads to marked rectification with forward to reverse current ratios of 10^2–10^4.

The preparation of *n*-type cuprous oxide is difficult; only certain impurities, which are absent in commercial rectifiers, make cuprous oxide *n*-type. However, Lashkarev has succeeded in finding *n*-type conduction at the boundary between the barrier layer and the copper.

An electromotive force is produced in cuprous oxide photocells (which differ from cuprous oxide rectifiers in that they have a semi-transparent outer electrode) when they are illuminated. If they are illuminated with red light, which penetrates the oxide layer, a photo-electromotive force is produced at the junction with the copper. This is the so-called back-contact photoeffect. Visible light of short wavelength, however, gives rise to a photoeffect at the front face, at the

junction of the oxide with the semi-transparent metal, provided that a barrier layer is formed between them.

The thermoelectric power α for cuprous oxide is about 7×10^{-4} V/deg at room temperature. Its temperature coefficient is small. Neither the magnitude nor the temperature variation of α is in agreement with theoretical calculations.

Among other oxides showing p-type conduction UO_2 has been studied. This has been used to make thermistors on account of its relatively high temperature coefficient of electrical conductivity and great stability.

Nickel oxide has been much investigated. Nickel atoms possess an unfilled $3d$ shell. Accordingly it would be expected that NiO would be a metal; however, it is in fact a typical semiconductor. The $3d$ shell does not usually take part in conduction, since the energy barrier for the electrons, as for electrons in deeper lying shells, is too large, while its penetration by the tunnel effect is inadequate.

The energy gap in nickel oxide is 1·7–1·9 eV. The thermoelectric power is inversely proportional to the absolute temperature.

Zinc oxide has been studied as both natural and synthetic single crystals, as sintered or compressed specimens, or in the form of a vacuum-evaporated layer.

The electrical conductivity of zinc oxide increases as oxygen is removed. σ increases with excess zinc concentration from 10^{-6} to 10 ohm^{-1} cm^{-1}, while its temperature coefficient falls. At high temperatures equilibrium is established with the oxygen in the surrounding medium. The relation between the electrical conductivity σ and the oxygen pressure p can be expressed by the formula

$$\sigma \propto p^{-1/4 \cdot 3}.$$

The energy gap is 2·6 eV, if it can, in fact, be estimated from the temperature coefficient of electrical conductivity at high temperatures under conditions when equilibrium with the surrounding medium is suppressed and the state of the material must be considered metastable.

The mobility of electrons in zinc oxide single crystals reaches 200 cm^2/V sec at room temperature, though in sintered specimens it does not exceed 50 cm^2/V sec. The mobility varies but little with temperature.

Manganese oxide has an energy gap of 1·25 eV. The dissociation energy of impurities varies from 0·03 eV for small concentrations to close to zero for large.

Titanium dioxide, TiO_2 is encountered in three modifications: anatase, brookite, and rutile. The last of these has the highest density

(4·22 gm/cm^3) and is the most stable. On heating to 1000° C anatase and brookite transform to rutile. Rutile is noteworthy for its extremely high dielectric constant: $\epsilon_{\parallel} = 167$ along a direction parallel to the principal axis, and $\epsilon_{\perp} = 83$ along a direction perpendicular to it.

The energy gap is 3·05 eV so that the intrinsic conductivity σ_0 at room temperature is small, less than 10^{-13} ohm^{-1} cm^{-1}; at 300° C σ_0 is 10^{-2} ohm^{-1} cm^{-1}, while at 1000° C σ_0 reaches 1000 ohm^{-1} cm^{-1}.

Rutile which has been partially reduced in hydrogen has a greenish colour which gradually becomes darker as more oxygen is removed and the composition approaches Ti_2O_3. At the same time as oxygen is removed the electrical conductivity due to electrons in the conduction band increases. For small concentrations the separation ΔE between the impurity levels and the conduction band is about 0·7 eV, which disagrees with the ionization energy of an outer electron from an oxygen atom in a medium with the dielectric constant of rutile. As the impurity concentration increases ΔE decreases, and at concentrations greater than 0·1% it approaches zero. The number of electrons becomes almost the same as the number of impurity atoms. At such concentrations the wave functions of the electrons of neighbouring impurity atoms overlap, and a continuous band is formed which leads to a metallic, impurity, conduction.

The mobility of electrons in rutile is about 1 cm^2/V sec at room temperature.

The optical absorption limit at room temperature is about 4050 Å, corresponding to 3·0 eV. On lowering the temperature to 20° K the limit moves to higher frequencies, corresponding to 3·2 eV; on increasing the temperature to 200° C the limit is displaced towards the red (2·8 eV). The position of maximum photoeffect also corresponds to the energy gap, if the latter is 3·0 eV at 290° K.

The thermoelectric power of titanium dioxide is of the order of 300 microvolts/deg, while the small magnitude of the Hall effect indicates a large concentration of electrons. It can be concluded from this that the effective mass of electrons in titanium dioxide is appreciably greater than the mass of the free electron.

High dielectric constants are possessed by the *n*-type semiconductors WO_3 and MoO_3. The Hall effect with these materials, as with TiO_2, is extremely small.

In spite of *n*-type conduction, which tends to lower the value of the contact potential, WO_3 is remarkable for its extremely high contact potential which exceeds that of gold. A junction between WO_3 and a *p*-type semiconductor does not give a rectifying barrier layer, since for this case, in conformity with the difference in contact potential, an anti-barrier layer with a reduced resistivity is formed.

Aluminium oxide, which is used as the dielectric in electrolytic capacitors, has been the subject of frequent investigations. The layer of oxide on the surface of the aluminium in such capacitors usually has a thickness of the order of 10^{-6} cm. Once such a layer has been formed, further oxidation of the aluminium is inhibited, the reason for this being the extremely low electronic conductivity of aluminium oxide which impedes the diffusion of positive aluminium ions to the external surface. At temperatures above 1000° C aluminium oxide begins to conduct and behaves as a typical semiconductor.

Oxides which have the spinel structure require special mention. Spinel is the name of the mineral $MgAl_2O_4$. Its structure can be described as a close-packed cubic lattice of negative oxygen ions within which are inserted positive metal ions. The sites of the metal ions, with respect to the oxygen ions surrounding them, can be of two types: part of them are situated in tetrahedral sites between four oxygen ions, while the remainder are in octahedral sites between six oxygen ions.

The unit cell of ordinary spinel is shown in Fig. 150. It consists of 32 oxygen ions; 8 tetrahedral sites are occupied by divalent metal ions, and 16 octahedral sites by tervalent ions of another metal. The 16 metal ions, occupying positions at the centres of octahedral spaces, form a continuous chain in which the distance between the ions is sufficiently small for transport of electrons by the tunnel effect to occur.

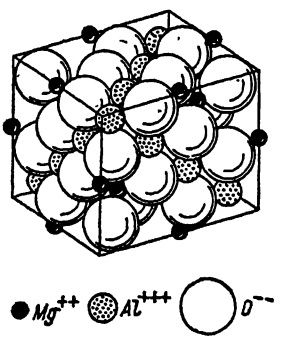


Fig. 150. Unit cell of normal spinel, $MgAl_2O_4$.

A range of compounds which crystallize in the spinel structure has recently become of great practical importance. Many of them are ferromagnetic materials with very high magnetic susceptibility and coercivity. Since such "ferrites", in contrast to the ferrous group of ferromagnetic metals, have a high electrical resistivity, they do not give rise to eddy-current losses even at frequencies of the order of several megacycles per second, where iron becomes useless from this cause.

Two types of spinel can be distinguished: normal and inverse. In an inverse spinel 8 tervalent ions occupy tetrahedral sites, and 16 octahedral sites are statistically distributed between 8 divalent and 8 tervalent ions, as, for example, in Fe_3O_4: [8 Fe^{+++} (8 Fe^{++}, 8 Fe^{+++}) 32O^{--}] (see Fig. 151). Ferrites are materials with inverse spinel lattices only. They form domains with similarly ordered spins (ferromagnetic materials) or with paired, oppositely ordered, spins (antiferromagnetic materials).

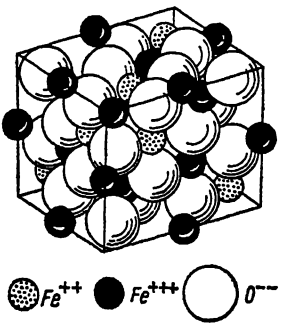


Fig. 151. Unit cell of the inverse spinel, Fe_3O_4.

Inverse spinels also show peculiarities in electrical behaviour: electron movement in them can take place not only in the conduction band, but also by means of transitions from one ion to another through the energy barrier between them.[55]

Divalent and tervalent ions of the same element can be present in the individual octahedral sites of an inverse spinel, and exchange of electrons can occur between them. In an electric field this statistical exchange becomes directional; preferred exchange takes place along those directions in which the electrical forces act.

The mechanism of current flow in magnetite, Fe_3O_4, is apparently of this type. In this material, as was noted above, one-third of the iron ions are divalent, and two-thirds tervalent, $Fe_3O_4 = Fe^{++}O^{--} + Fe_2^{+++}O_3^{--}$. By transferring from a divalent to a tervalent ion, an electron interchanges their sites, while an electron from another divalent ion can then move to the vacated position, and so on.

In normal spinels this type of charge transfer is not possible, since di- and tervalent ions are on different crystallographic sites and their electrons are of different energy. Each electron transfer requires an overall expenditure of energy. In inverse spinels, however, the overall result is the interchange of ions in positions of the same type without any change in total energy.

Oxides with structures of the perovskite type are often ferroelectric. To this group belong the titanates of the second group, which were first discovered by Vul, and in particular the ferroelectric material, barium titanate, $BaTiO_3$, the dielectric constant of which can reach 3000, while its Curie temperature lies above 100° C. With Seignette salt, the special properties of which were discovered previously by Kurchatov and Kobeko, the Curie temperature is 24° C.

Both barium titanate and Seignette salt have found wide practical application as piezoelectric materials in sound-reproducing apparatus.

Ionic crystals

Rock salt, NaCl, is a typical example of an ionic crystal. Its cubic crystals are built up by the alternate positioning of positive sodium ions and negative chlorine ions along the cube edge. The electrical attraction between the sodium and chlorine ions holds the lattice together.

The electrical conductivity of such crystals is predominantly ionic.. However, under the influence of irradiation by X-rays or an electron beam, individual light-absorbing centres appear in the crystal. These give rise to electronic photoconduction when illuminated with visible light. This effect was first observed and studied in 1903.

In 1921 Röntgen published a detailed quantitative description of the properties of natural and X-irradiated rock salt which was based on simultaneous experiments by him and by us.

Before this, in 1916, Lukirskii showed, by means of the Hall effect, that the photocurrent carriers in rock salt are electrons.

At about this time (beginning in 1920) the photoeffect in alkali halide crystals was studied by Pohl and his co-workers. He showed that photosensitive salt could be obtained not only by subjecting it to the action of ionizing radiations but also by heating it in an atmosphere of sodium vapour.

As the result of 30 years, or perhaps more correctly 50 years, of study, it has been found that the photosensitivity of salt is due to the so-called *F*-centres. These are lattice sites at which a negative chlorine ion is replaced by an electron. The case of two electrons at such a site has been termed an *F'*-centre.

A most complete and perfect theory of *F*- and *F'*-centres was developed by Pekar in connection with his theory of polarons.[56, 57] Extensive experimental data on *F*- and *F'*-centres are presented in Mott and Gurney's book *Electronic Processes in Ionic Crystals*. My own experiments for the years 1903–1906 are described in a monograph, *The Physics of Crystals*.

In silver chloride electrolytic conduction occurs by means of silver ions, while the appearance of electronic conduction on illumination is

due to electrons and holes. The mobility of electrons at room temperature is 70 cm^2/V sec, while the hole mobility is half this value. The lifetime of charge carriers is large.

There exists a large number of crystals with predominantly ionic bonding in which not only the photoconductivity but also the dark conductivity is electronic, or partially ionic and partially electronic. Many oxides and sulphides have ionic or predominantly ionic bonding.

All semiconductors can be assigned to a continuous sequence, which is characterized by a gradual transition from compounds with purely ionic bonding to compounds with purely covalent bonding. In addition to foreign atoms introduced into the crystal, excess or deficit of one of the constituents can act as a donor or acceptor of electrons. Excess atoms can be situated between the lattice sites, distorting the lattice in their vicinity. A similar result is obtained by the removal from the lattice of ions of opposite sign, leaving behind unoccupied sites. In both cases the region in which distortion of the lattice takes place can diffuse under the influence of thermal motion, and in an electric field, is preferentially displaced along the direction of the field constituting an electrolytic, ionic, current.

Electronic conduction is not observed in pure ionic crystals in which the energy gap is so large by comparison with the dissociation energy of ions in the lattice that the numerical predominance of ionic charges outweighs the marked difference in mobilities. The availability of centres with electronic levels situated within the forbidden gap changes the situation, since the transfer of electrons from these levels requires less energy and can occur more frequently. In such crystals electronic conduction is found in addition to ionic conduction, and in some cases can appreciably exceed it.

The dielectric displacement of ions in an electric field produces an appreciable polarization in the regions of the crystal surrounding the free charge and lowers its energy. Pekar termed such a state of free charge a polaron and showed that the transport of current in ionic crystals takes place largely by polarons.

Alloys of metals

Kurnakov and his co-workers observed, on melting-point curves of alloys, sharp maxima which occurred at definite stoichiometric ratios of the constituents. Kurnakov termed alloys of such composition daltonide (Fig. 152), while alloys corresponding to flatter maxima he termed berthollide (Fig. 153).

Until recently it was not clear whether or not such alloys should be considered as metals. Investigations of their electrical properties over a wide range of temperatures by Zhuze, Boltaks, Mochan, and others,

has shown that they are semiconductors, possessing an energy gap.[58] For example, measurements of the Hall effect and electrical conductivity

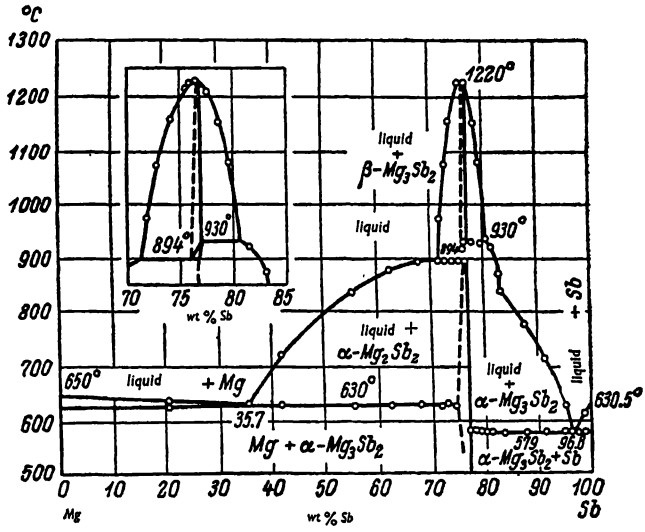


Fig. 152. Phase diagram of the system Mg-Sb. Mg_3Sb_2 is a valence intermetallic compound of daltonide type.

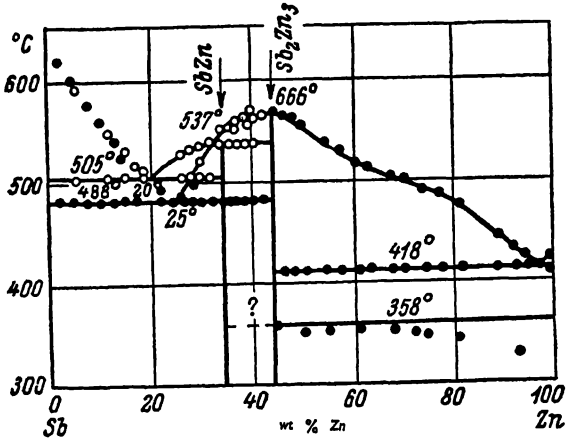


Fig. 153. Phase diagram of the system Zn-Sb. ZnSb is an intermetallic compound of berthollide type.

of the alloy Mg_3Sb_2 have shown that it behaves as a semiconductor with an energy gap of 0·7 eV and with impurity levels approximately 0·1 eV from the appropriate band edge.

Mochan, using a method developed by Vekshinskiĭ[59] (employing the simultaneous deposition, by evaporation from different directions in a vacuum, of both metals) obtained layers of alloys of Mg-Sb with continuously varying composition. The region corresponding to the daltonide compound Mg_3Sb_2 was a poor conductor with a resistivity of the order of 10^4 ohm cm, a thermal conductivity less than 0·01 cal/cm deg sec, and a transparency to visible light. This region also showed marked photoconduction on either side of this stoichiometric region, being *n*-type on the side with excess magnesium and *p*-type on the side with excess antimony.[60]

A rather different type of semiconducting alloy is the berthollide compound ZnSb, which has been studied in detail by Maslakovets and Devyatkova. In this case the highest resistivity is reached with equal atomic proportions of zinc and antimony, which is not in agreement with their valencies. Excess of both zinc and antimony leads to increased carrier concentrations, but in both cases the alloy shows *p*-type conduction.[61]

A semiconductor-like variation of electrical conductivity and Hall constant has been found with a number of other alloys: Mg_2Sn, CdSb, Cs_3Sb, etc. From the temperature variation of the Hall constant over a wide range of temperature one can determine both the energy gap ΔE_0 and the ionization energy ΔE of levels produced by excess or deficiency of the constituents of the alloy with respect to the stoichiometric composition.

Metal alloys so far studied which have an energy gap have the following characteristics:

	ΔE_0 (eV)
Cs_3Sb	1·0
Mg_3Sb_2	0·7
ZnSb	0·4
CdSb	0·52
Mg_2Si	0·77
Mg_2Ge	0·74
Mg_2Sn	0·36

The temperature variation of the mobility for the last three alloys has the form $u \propto T^{-\frac{1}{2}}$.

The highest electronic mobility among these alloys has been found with CdSb, for which it can reach 660 cm^2/V sec. Excess of Sb converts CdSb to *n*-type, excess of Cd to *p*-type. Rectification is observed at a *p*–*n* junction.

In the alloy ZnSb, which is analogous to CdSb, *p*-type conduction is obtained with excess of both Zn and Sb. The hole mobility is about 200 cm^2/V sec at room temperature.

Liquid semiconductors

The very fact of the existence of liquid electronic semiconductors excludes the possibility of ascribing the discrete structure of the energy spectrum to strict periodicity (long-range order) of a crystal lattice. This fact was established by Regel' with Tl_2S and other materials, which show semiconducting properties right up to their melting-point, and retain these properties without any marked quantitative changes even after melting. In any case the long-established electronic conductivity of liquid solutions of sodium and potassium in ammonia indicates the possibility of the existence of electronic conductivity in the amorphous state. Tungsten bronzes, which are solutions of sodium in solid tungsten trioxide, are not very different in their properties from liquid solutions of Na in ammonia.

The transition from solid to liquid states is accompanied by a jump in the electrical conductivity, which has a negative or positive sign depending on whether the specific volume increases or decreases on melting. The temperature dependence of the electrical conductivity also increases in the one case and decreases in the other.

Some semiconductors, such as germanium, for example, lose their semiconducting properties on melting and become metals; others, such as selenium, tend to become insulators.

An analysis of all the experimental data accumulated by Regel' and his co-workers[62] leads to the belief that the decisive factor in the determination of electrical properties is not long-range but short-range order, i.e. the arrangement of nearest neighbours about a given atom or molecule. Thus germanium, which in the crystalline state has a coordination number (number of nearest neighbours) of 4, goes over into a more closely packed structure with 8 neighbours on melting, i.e. to a structure characteristic of many metals, for example, Na, K. In agreement with such a rearrangement in the short-range order the specific volume of germanium decreases on melting (and the density increases) (Fig. 154), while the electrical conductivity becomes metallic, increasing by more than an order of magnitude (Fig. 155). In those cases in which the short-range order, the number and position of nearest neighbours, remains unchanged on melting, as in the case of Te for example, the material retains its semiconducting properties.

With the transformation of hexagonal grey selenium into the liquid state, which is accompanied by an increase in specific volume of almost 13%, an energy gap is retained, but it increases in width, and the electrical conductivity of selenium decreases by several orders of magnitude.

With tellurium the change in short-range order occurs gradually with further heating of the liquid phase. This leads to a corresponding alteration in electrical properties, also.

It can be taken as established that semiconducting properties are not confined to the crystalline state, but can also be present in the amorphous state. So far, however, no theory of amorphous semiconductors has been developed. The quantum theory of semiconductors, with its band structure of energy levels, is linked exclusively to long-range crystal order.

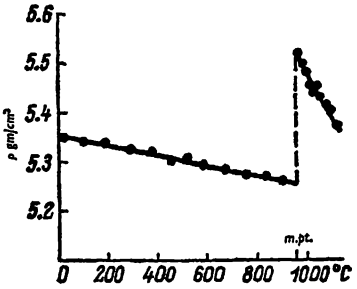


Fig. 154. Temperature dependence of the density for germanium.

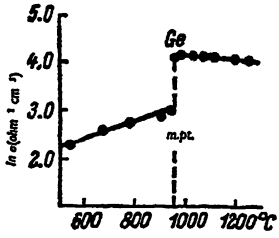


Fig. 155. Temperature dependence of the electrical conductivity of germanium.

Organic materials

Semiconducting properties were established by Vartanyan for layers of certain organic dyes. Energy gaps were determined by him for a number of materials from the slope of the plot of $\log \sigma$ against $1/T$. The data he obtained are presented below.[68]

	ΔE_0 (eV)
Phthalocyanine . . .	0·87
,, (sulphurized) .	0·9
,, (copper) . .	1·2
,, (magnesium) .	0·78
Methylene blue . . .	1·05
Crystal violet . . .	0·74

The photoeffect in dye films grows slowly and then falls off over a period of 10 to 30 minutes. The maximum photoeffect is observed at absorption bands, both in the visible and in the ultra-violet region of the spectrum.

Not only the dark current, but also the additional photocurrent is well represented by the dependence $\ln J = A - B/T$ with the same slope, approximately for both cases.

Similar behaviour was found by Zhuze and Ryvkin for the photocurrent in cuprous oxide. It may be supposed that the photocurrent in organic dyes is of secondary origin: the absorbed photons transfer their energy to excitons, some of which on encountering impurities become broken down to yield an electron or a photon. For most of the dyes investigated the current carriers are electrons.

Other organic materials, for example, anthracene, coronene, etc., also possess semiconducting properties.

Absorbed gases (oxygen and hydrogen) exert a marked influence both on the dark conductivity and on the photosensitivity. The effects show hysteresis and depend on the previous history of the layer.

The additional conductivity arising as a result of absorption of oxygen is proportional to the oxygen pressure.

Semiconductor materials for thermoelements

In every application of semiconductors, special requirements have to be met in the selection of the materials. Thus *p–n* junctions require a high mobility, a long free carrier lifetime and a large forbidden gap width. Germanium, silicon, silicon carbide, and their analogues are of special interest as materials for these applications. Materials for thermistors must have a high temperature coefficient of the electrical conductivity, mixtures of oxides or sulphides being most frequently used in such devices. Photoelectric batteries, utilizing solar energy, are made of materials which have to satisfy certain conditions with regard to the forbidden gap width to conform with the prevalent part of the solar spectrum; silicon and cadmium telluride are used for this purpose. The spectral curves of selenium photocells resemble closely the sensitivity of the human eye.

The section on thermoelements contains a discussion, on the basis of which it is possible to assess the suitability of various semiconductors for this application. We have to deal here with materials having a high electrical conductivity, of the order of 10^3 $\text{ohm}^{-1}\,\text{cm}^{-1}$, whereas materials for *p–n* junctions have an electrical conductivity of the order of 0·01 $\text{ohm}^{-1}\,\text{cm}^{-1}$ or less, whilst materials for photocells have an even lower electrical conductivity. The requirements of a low thermal conductivity and a high mobility, which materials for thermoelements have

to satisfy, focus our attention on elements in the lower part of the periodic system and on compounds which do not contain strongly ionic bonds.

Seebeck's table of thermoelectric materials started with lead sulphide and ended with the 50% zinc–50% antimony compound. A couple between these two materials has, in fact, good properties. Replacement of lead sulphide by lead telluride and the addition of tin to the zinc–antimony compound appreciably improves the quality of the thermoelements. Further progress was achieved by the use of bismuth telluride. The next step in improving the figure of merit z consisted of the use of the solid solutions $PbTe + PbSe$, $Bi_2Te_3 + Bi_2Se_3$, and $Bi_2Te_3 + Sb_2Te_3$ and the addition of iodine. At present it is possible to attain in this way values of z up to $3{\cdot}3 \times 10^{-3}$ deg^{-1}.

It will readily be seen from the expressions for the efficiency of thermoelectric generators and refrigerators that the figure of merit z always appears in a product with the term $(T_1 + T_2)/2$, i.e. with the mean temperature T_{mean} of the thermoelement, and the temperature T appears in the form T_1/T_2. Whilst z has the dimension deg^{-1}, the product zT_{mean} is a dimensionless quantity of the order of unity and the same applies to T_1/T_2.

Since generators usually work at cold junction temperatures only slightly in excess of room temperature, i.e. around 300–320° K, whilst in refrigerating devices, it is usually necessary to cool down from room temperature, the generator efficiency is governed by the hot junction temperature of the thermoelements, and the coefficient of performance of a refrigerator by the cold junction temperature.

Out of the two terms in the expression for the efficiency of a generator, the thermodynamic multiplier $(1 - T_2/T_1)$ is of the greater significance. Therefore, attention should be focused on raising the temperature T_1, at which operation of the thermoelement is still possible, even if this leads to a drop in z. When zT_{mean} does not vary appreciably, the efficiency increases with T_1 in the same way as does the thermodynamic multiplier. Raising T_1 leads to various difficulties: the diffusion of impurities is accelerated, oxidation occurs, cracks are caused by the large temperature gradient and the contact with the metal bridge joining the branches of the thermoelement deteriorates. The period during which the thermoelement retains its favourable properties is reduced.

We shall describe the materials which have been used until now for the construction of thermoelements. Seebeck used two compounds of zinc with antimony, Zn_3Sb_2 and ZnSb, and preferred the latter. The equilibrium diagram of this system shows the presence of a few other compounds such as Zn_3Sb_2 and Zn_4Sb_3.

In contrast with other semiconducting compounds, departures from the stoichiometric composition of ZnSb to either side do not lead to other than hole type of conduction, although they appreciably affect the value of the electrical conductivity, the thermal conductivity, and the thermoelectric power. The thermoelectric power reaches a sharp maximum of 280 μV/deg at 35 wt. % zinc, which corresponds to ZnSb.

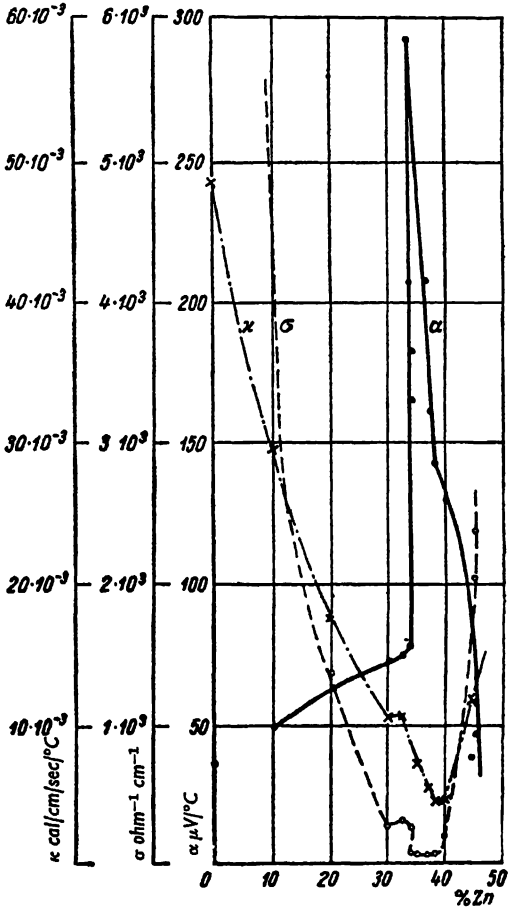


Fig. 156. Thermoelectric properties in the zinc–antimony system.

The electrical conductivity and thermal conductivity are then at their minimum values (Fig. 156). Specimens with less than 35% Zn contain, in addition to ZnSb, an excess of Sb, whilst, in specimens with a higher Zn content, we find ZnSb together with Zn_3Sb_2 and Zn_4Sb_3 (Fig. 156).

The lowest thermal conductivity for ZnSb is approximately $4{\cdot}5 \times 10^{-3}$ cal/deg cm sec, the electrical conductivity approximately

75 $ohm^{-1}\,cm^{-1}$, and $\varkappa$ reaches values of 1×10^{-3} to $1{\cdot}2\times10^{-3}\,deg^{-1}$. Figs. 157 and 158 show the temperature dependence of the hole concentration n_+ and the mobility u_+. Compounds with an excess of 1% Zn give the best results as materials for the positive branch of thermocouples. Impurities consisting of elements in columns I, IV and VI of the periodic table increase the hole concentration, whilst impurities consisting of elements of columns II, III and V reduce it. Of the

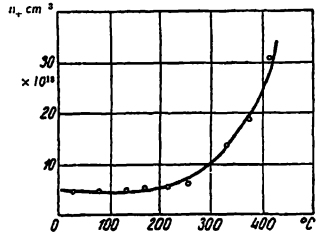


Fig. 157. Hole concentration against temperature in ZnSb.

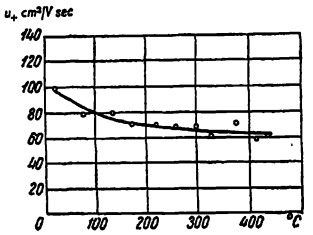


Fig. 158. Hole mobility against temperature in ZnSb.

elements of the fourth column, the addition of 0·5% Sn increases the electrical conductivity of ZnSb to more than double the value for the pure compound. The mobility is then at its maximum and falls slowly when the Sn content is further increased. The mobility varies little with temperature. The part of the mobility which depends on scattering by phonons varies as $T^{\frac{1}{2}}$. The value of $\varkappa$ for ZnSb with 1% Zn and 0·5% Sn as impurities reaches $1{\cdot}5\times10^{-3}\,deg^{-1}$ at room temperature.

Lead telluride and lead selenide, as well as solid solutions of these compounds, are good materials for the negative branch of thermoelements. Their parameters are better than those of the lead sulphide which was at Seebeck's disposal. Some data on PbSe and PbTe will be reported below. On the basis of results obtained on insufficiently pure specimens, it was considered that, on departures from the stoichiometric composition, an excess of lead sets up electronic conduction, and an excess of tellurium hole conduction. However, Koval'chik and

Maslakovets showed that, for chemically pure PbTe, the introduction of excess Pb has an effect only when the amount added does not exceed 1×10^{19} atoms per cm^3, reducing the electrical conductivity. Only after the addition of bromine, chlorine, or iodine does lead become a donor. Tellurium by itself is similarly ineffective.

According to these authors, a divalent tellurium atom may be replaced by two bromine atoms with two valence electrons, an electropositive site being left vacant. If the vacant site is now occupied by a lead atom, the latter contributes two electrons to the conduction band, i.e. one electron for each bromine atom introduced.

Cl, Br, I, Bi, Sb and Nb impurities are donors; Ag and Ta are acceptors. A range of metals possesses similar properties to excess lead; these metals include Sn, Ge, Ti, Ni, Co, Fe, Pt and Mg. Finally, Se, S and O behave like excess Te.

The activation energy of the charge carriers, or the width of the forbidden gap, calculated from the temperature dependence of the electron concentration in quenched specimens, amounts to 0·62 eV, which is in accordance with the heat of formation of PbTe from the two elements. However, optical measurements indicate that ΔE_0 is of the order of 0·35 eV. It is possible that this discrepancy is due to the change of the composition of PbTe with temperature; as a result of the solution of excess tellurium, the rate of rise of concentration with temperature increases. However, the forbidden gap width at all these temperatures still remains close to 0·35 eV.

The concentration of free electrons may be varied from 6×10^{17} to 2×10^{20} cm^{-3} and the concentration of holes from 4×10^{18} to 2×10^{19} cm^{-3}. Above room temperature, the electron mobility decreases with temperature, in PbTe, almost in proportion to T^{-3} and, in PbSe, in proportion to $T^{-2.5}$. At temperatures below 200° K the mobility decreases at a slower rate, namely as $T^{-1.5}$.

At room temperature the mobility of PbTe is about 1500 cm^2/V sec, that of PbSe is 1200 cm^2/V sec, and that of their solid solutions about 900 cm^2/V sec. The variation of mobility with concentration of PbSe in the solid solution and at different absolute temperatures is shown in Fig. 159.

The thermal conductivity is about 5×10^{-3} cal/deg cm sec for PbTe and PbSe and only to $2{\cdot}2 \times 10^{-3}$ cal/deg cm sec for solid solutions.

The thermoelectric power increases with increasing temperature. Its values agree with the values calculated from Pisarenko's expression on the assumption that, either

$$r = 0, \quad m_{\perp}^* = 0{\cdot}325 m_0,$$

or

$$r = -1, \quad m_{\perp}^* = m_0.$$

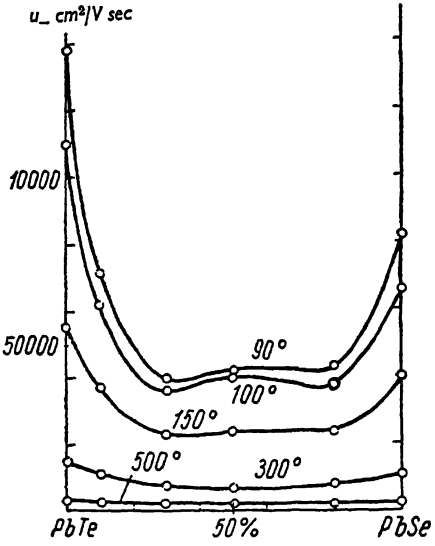


Fig. 159. Mobility of electrons in PbTe–PbSe alloys. (Temperatures in degrees absolute).

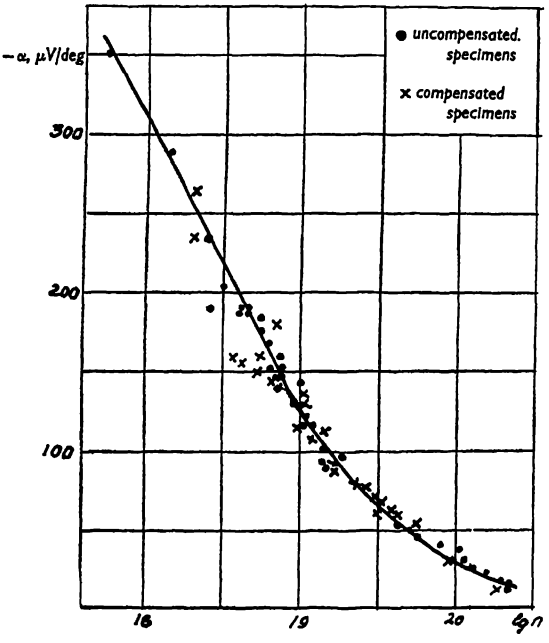


Fig. 160. Dependence of the thermoelectric power on the electron concentration in PbTe.

For various reasons it is the first of these assumptions that has to be adopted. The dependence of the thermoelectric power on the concentration n (Fig. 160) indicates the onset of degeneracy at 2×10^{19} electrons per cm^3. The dependence of α on the absolute temperature for concentrations n from $5{\cdot}7 \times 10^{17}$ to $1{\cdot}8 \times 10^{20}$/cm^3 is shown in Fig. 161.

Bi_2Te_3 and Sb_2Te_3 must be regarded as the best compounds, at present in use, for thermoelements working at temperatures not exceeding 250° C and for all cooling applications. The crystals of both substances belong to the rhombohedral system and are readily cleaved along planes perpendicular to the principal axis. The width of the forbidden gap in Bi_2Te_3 is, at room temperature, 0·16 eV; in Sb_2Te_3 the bands appear to overlap slightly.

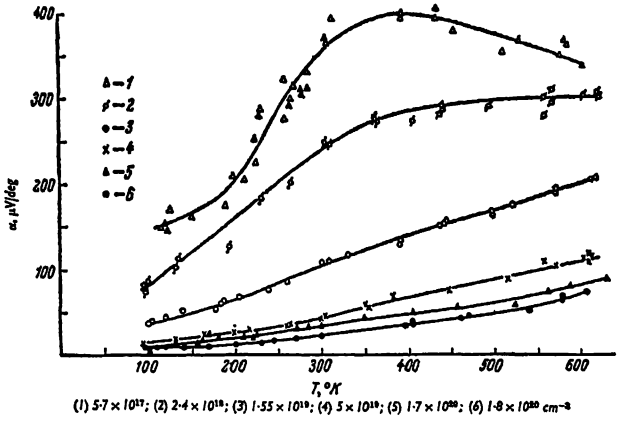


Fig. 161. Variation of thermoelectric power with temperature for samples of PbTe.

Investigation of the galvanomagnetic properties of Bi_2Te_3 led Drabble and Wolfe to the adoption of an energy band model in quasi-momentum space consisting of six separate bands with energy minima at reflection planes. The effective mass ratios in each band are $m_1/m_2 = 1{\cdot}21$ and $m_3/m_2 = 0{\cdot}093$.

The effective mass of electrons travelling along the cleavage plane is found to be $1{\cdot}3m_0$ to $1{\cdot}9m_0$ if the anisotropy of the Hall effect is not taken into account.

The ratio of the resistivity $\rho_\perp$, in the direction perpendicular to the cleavage plane, to the resistivity $\rho_\parallel$, along the plane in which the crystals are usually studied and used, is approximately equal to 4. The thermoelectric power α, which reaches up to 250 μV/deg, both in p-type and n-type crystals, does not depend on the crystallographic orientation.

The thermal conductivity $\kappa_{\parallel}$ is somewhat higher than $\kappa_{\perp}$. The pronounced laminar structure of the crystals, the configuration of atoms and the distance between the atoms in the lattice led Airapetyants and Efimova to the configuration of Bi, or Sb, atoms and Te atoms depicted in Fig. 162. It will be noted that, of the Te atoms, one layer Te_I is arranged between two Bi atom layers, whereas Te_{II} atoms are between

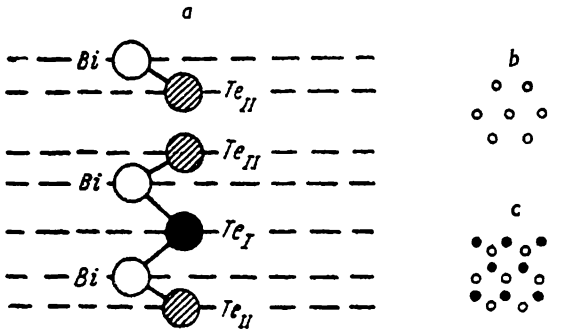


Fig. 162. Disposition of atoms in Bi_2Te_3.

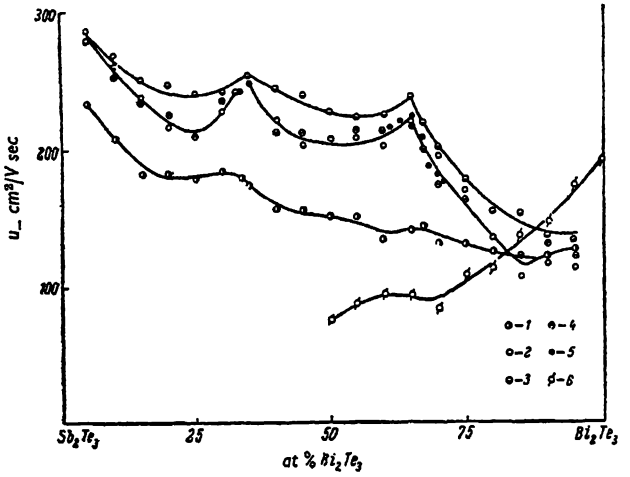


Fig. 163. Mobility of Bi_2Te_3–Sb_2Te_3 alloys.

layers of Bi and Te_{II}. In solid solutions of $Bi_2Te_3 + Sb_2Te_3$ additional atoms may be distributed either statistically, or (as a result of annealing) become ordered when the proportions of Sb and Bi amount to $\frac{1}{3}$ and $\frac{2}{3}$ or $\frac{1}{2}$, $\frac{1}{4}$ and $\frac{3}{4}$. The mobility is higher in ordered solid solutions (Fig. 163). In Bi_2Te_3 the binding between atoms within each Bi layer and within each Te layer is weak, being primarily of the Van der Waals

type. The binding between atoms in the two adjacent Te_{II} tellurium layers is also weak. Te_I atoms, which lie between two bismuth layers, are joined to the bismuth layers by strong ionic binding.

In Sb_2Te_3 the adjacent tellurium layers, Te_{II}, are bound to each other with a stronger bond and are bound to neighbouring Sb atoms by covalent bonds. The binding between the tellurium atoms Te_I and the two adjacent antimony layers has a less pronounced ionic character.

Sinani and Gordyakova established that iodine, copper and tellurium impurities act in Bi_2Te_3 and Sb_2Te_3 as donors, whereas bismuth, antimony, lead and tin act as acceptors. Iron, sodium, chlorine and some other impurities do not dissolve in appreciable quantities and therefore do not affect the concentration of charge carriers.

In Bi_2Te_3 each CuBr molecule yields one free electron, whilst three lead atoms yield one hole.

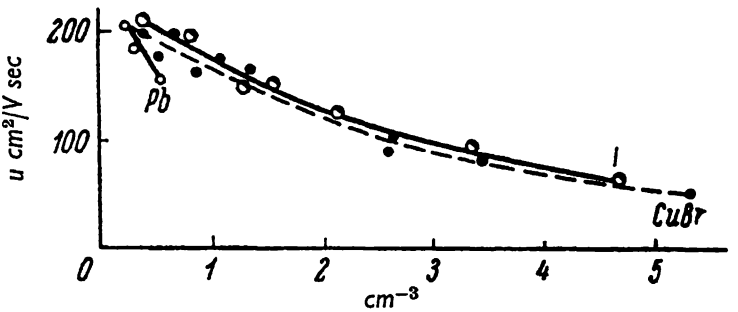


Fig. 164. Effect of impurities on the mobility of Bi_2Te_3.

The effect of I, CuBr and Pb impurities on the mobility u may be judged from Fig. 164, where n denotes the number of charge carriers arising from the impurities.

Excess of Bi and Te above the stoichiometric composition of Bi_2Te_3 leads to an increase in the concentration of holes when Bi is in excess, and of electrons when Te is in excess; this may be seen from Fig. 165, published by Satterthwaite and Ure. The dependence of the thermoelectric power α on the proportions of Bi and Te is shown in Fig. 166, also due to these authors.

The highest purity Bi_2Te_3 specimens contained approximately 10^{15}–10^{17} electrons or 10^{18} holes.

Goldsmid investigated in detail the electrical and thermal properties of Bi_2Te_3. He studied the effect of degeneracy on the mobility, the electronic thermal conductivity due to both free carriers arising from impurities and, also, the diffusion of electron–hole pairs, as well as the anisotropy of the electrical and thermal properties.

The quality of thermoelectric materials improves with increasing ratio of mobility to thermal conductivity.

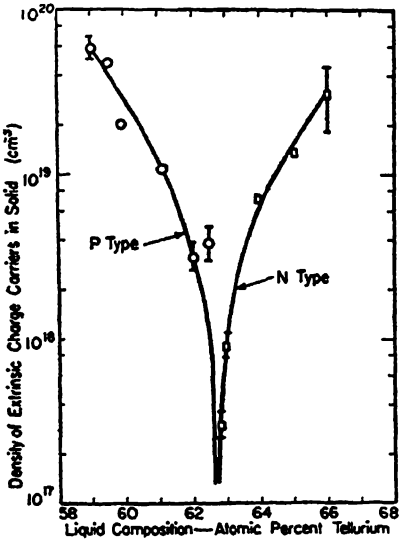


Fig. 165. Carrier density in Bi_2Te_3 of non-stoichiometric composition.

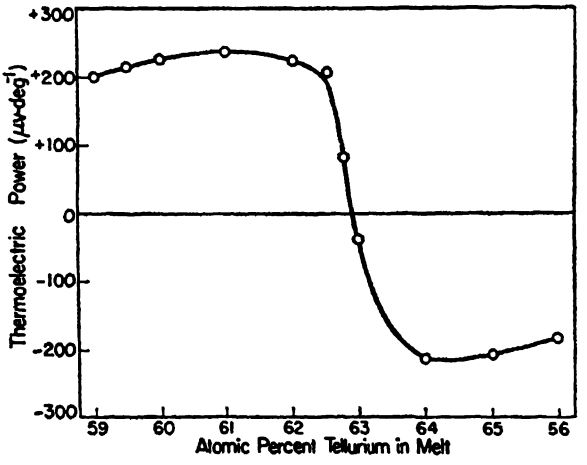


Fig. 166. Effect of composition on the thermoelectric power of Bi_2Te_3.

The thermal conductivity may be reduced by the introduction of impurities, which act as additional phonon scattering centres. A. V. Ioffe showed that this reduction is particularly pronounced in solid solutions in which the periodic structure is affected so that there is scattering of

phonons with wavelengths of the order of inter-atomic distances. Examples of such systems are discussed in the section on thermal conductivity. At the same time the scattering by individual impurity atoms of electrons with wavelengths of the order of 20 interatomic distances should be less pronounced. These considerations were experimentally substantiated and solid solutions were found to be more suitable materials for thermoelements than either of the constituents in the pure form.

However, the scattering of electrons is also characterized by another important feature. In contrast to phonons, the waves of travelling free carriers are charged electrically and become denser close to ions with the opposite charge and less dense close to ions with a similar charge. Therefore, when in the crystalline lattice the arrangement of ions of one sign is impaired, for example, that of positive ions, this will lead to increased scattering of the charge carriers of the opposite sign, in this case electrons. The disturbance of the arrangement in the negative ion sublattice leads to increased scattering of holes.

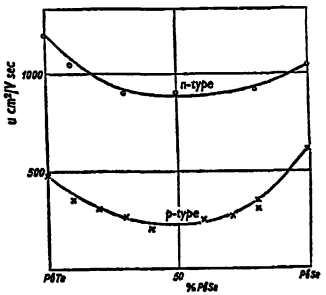


Fig. 167. Carrier mobility in PbTe–PbSe solid solutions.

Figs. 167 to 170 illustrate the properties of solid solutions described above.

Fig. 167 shows the scattering of electrons and holes in PbTe–PbSe solid solutions at different concentrations. The Pb^{++} positive ion sublattice here is unchanged, whereas a statistical spread is introduced into the system of negative ions Te^{--}. As a result of this, the electron mobility is reduced by only 20%, whereas the hole mobility decreases by a factor of 2·5.

Fig. 168 makes it possible to compare a reduction of the thermal conductivity in the same solid solutions by a factor of about 3 with the effect on the electron mobility.

Fig. 169 enables a comparison to be made between the effects on the mobility in Bi_2Te_3 and Bi_2S_3 solid solutions due to the retention of the

periodicity of the positive bismuth ion sublattice and the disturbance of the arrangement of the negative tellurium ions. The electron mobility remains virtually the same up to concentrations of 30% Bi_2S_3 whereas the hole mobility decreases by a factor of 4·5.

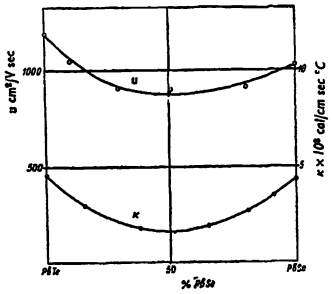


Fig. 168. Electron mobility and thermal conductivity in PbTe–PbSe solid solutions.

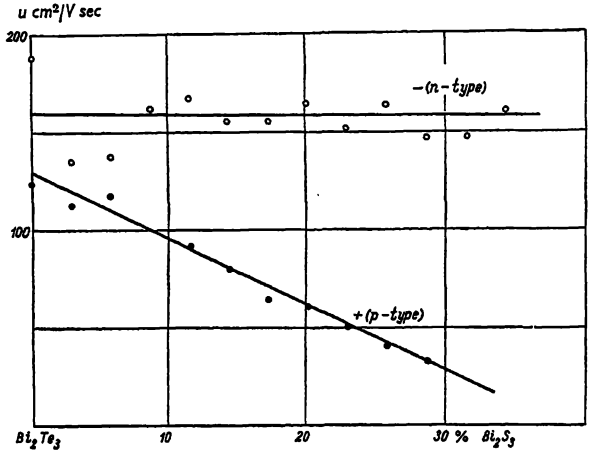


Fig. 169. Carrier mobility in Bi_2Te_3–Bi_2S_3 solid solutions.

Fig. 170 shows that, where the periodicity of positive Bi and Sb ions is upset, the electron mobility decreases to a half, whereas the hole mobility increases by a factor of 1·7 as a result of the ordering of the solid solution, which reaches a maximum at 33% Sb_2Te_3.

When introducing impurities, in order to provide the optimum concentration of free charge carriers, it is necessary to bear in mind the sublattice which will be distorted by a given impurity. One should introduce into *n*-type materials such impurities as will upset the arrangement of negative ions. On the other hand, in *p*-type semiconductors,

the impurities should only upset the periodicity in the positive ion sublattice. If this rule is not followed, the impurities reduce the mobility of free charge carriers to too great an extent.

Several investigations of solid solutions have recently been reported in connection with the aforementioned hypothesis of A. V. Ioffe. Thus Sinani, Gordyakova and Kokosh studied the properties of Bi_2Te_3–Bi_2Se_3 solid solutions and Bi_2Te_3–Sb_2Te_3 solid solutions with the compositions of 80% Bi_2Te_3–20% Bi_2Se_3 and 74% Sb_2Te_3–26% Bi_2Te_3, which according to X-ray analysis give single-phase systems. The forbidden gap width for the 25% Bi_2Te_3 and 75% Sb_2Te_3 solution was found to be equal to 0·1 eV.

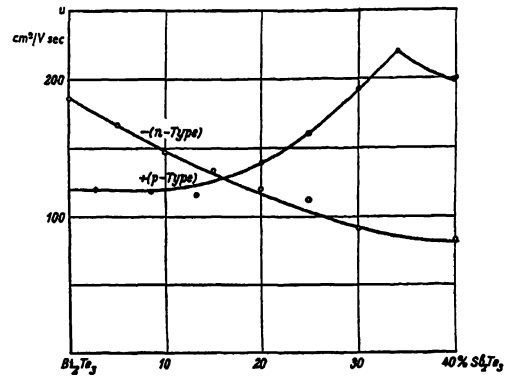


Fig. 170. Mobilities of electrons and holes in Bi_2Te_3–Sb_2Te_3 solid solutions.

Solutions containing 15–25% Bi_2Se_3 and 85–75% Bi_2Te_3 doped with 0·08% CuBr had the best properties for the *n*-type thermoelement branch, yielding z of up to $3{\cdot}3 \times 10^{-3}$ deg^{-1}. Iodine and copper as impurities yield one electron for each impurity atom.

When the situation is such that the scattering of electrons by the impurities predominates over the scattering by phonons, then the thermoelectric power

$$\alpha = \frac{k}{e}\left(r + 2 - \frac{\mu}{kT}\right)$$

should increase owing to the increase of r. The excess electron concentration may be reduced by introducing, together with the iodine donors, lead acceptors. However, no effect on α could be observed at room temperature.

A solid solution of 74 mol % Sb_2Te_3 with 26% Bi_2Te_3 for the positive branch had the parameters $\alpha = 155$ μV/deg, $\sigma = 1500$ ohm^{-1} cm^{-1}, $\kappa = 4{\cdot}3 \times 10^{-3}$ cal/deg cm sec and $z = (2{\cdot}3$ to $2{\cdot}6) \times 10^{-3}$ deg^{-1}. The

addition of 0·3% and 3% Te, which only partly dissolved and was mainly in the form of a eutectic, and of 0·1% Pb or 0·08% I led to an increase of z to (3·2 to 3·3) $\times 10^{-3}$ deg^{-1}.

In order to produce each extra hole in a compound containing 3% Te, it was necessary to add 3·3 Cu, 0·8 Ag, 1·4 Pb or 1·4 Sn atoms. Iron did not enter into the solid solution at all.

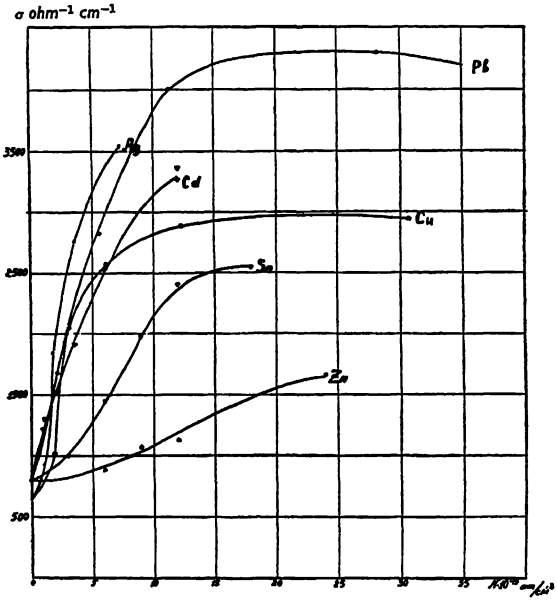


Fig. 171. Dependence of electrical conductivity on impurity concentration for 74% Sb_2Te_3–26% Bi_2Te_3 with 0·3% Te excess.

The relationships between σ and α and N and between $\alpha^2\sigma$ and σ, for an excess of 0·3% Te in the solid solution, are shown in Figs. 171, 172 and 173 respectively, and the dependence of $\alpha^2\sigma$ on σ for the solid solutions containing 0·3% and 3% Te and $PbCl_2$ is shown in Fig. 174.

As has been mentioned earlier, an increase of the hot junction temperature is of particular importance for raising the efficiency of thermoelectric generators. All the aforementioned materials, ZnSb, PbTe, PbSe and their solid solutions, Bi_2Te_3, Sb_2Te_3 and alloys of Bi_2Te_3 with Bi_2Se_3 and of Bi_2Te_3 with Sb_2Te_3, cease to be satisfactory at temperatures above 400° C. Therefore persistent attempts are being made to find semiconductors which retain their good properties up to 700–800° C. The best results among the materials investigated up to now are exhibited by metallic alloys with semiconducting properties: $CoSb_3$, Mn_2Si, CrSi and Si + Ge.

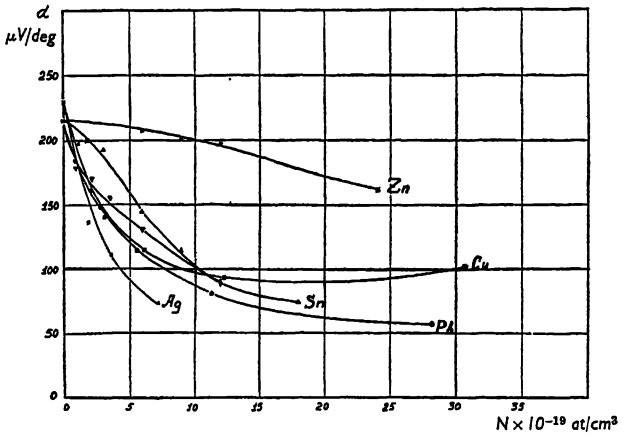


Fig. 172. Dependence of thermoelectric power on impurity concentration for 74% Sb_2Te_3–26% Bi_2Te_3 with 0·3% Te excess.

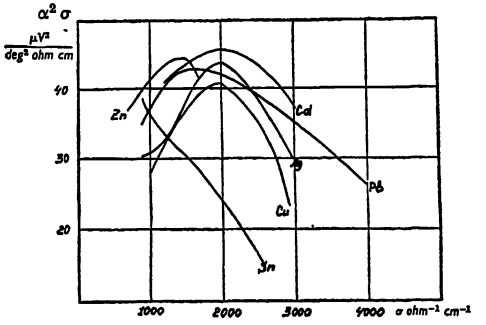


Fig. 173. Dependence of $\alpha^2\sigma$ on σ for 74% Sb_2Te_3–26% Bi_2Te_3 with 0·3% excess Te and various additional impurities.

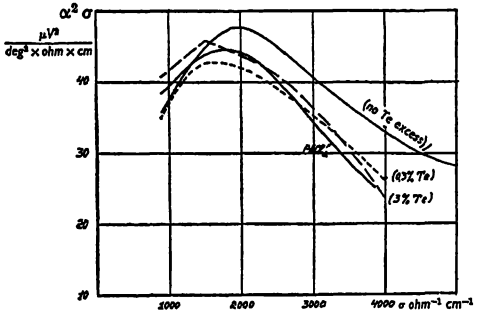


Fig. 174. $\alpha^2\sigma$ plotted against σ for 74% Sb_2Te_3–26% Bi_2Te_3 with 0·3% and 3% excess Te and with $PbCl_2$.

The general physical condition for all alloys with a degenerate system of free charge carriers is that the Fermi level should be such that the density of states in the electron energy level system is as great as possible, whilst α should reach values of 100 μV deg^{-1} and more.

Whereas materials are available for moderate temperatures with z equal to $3{\cdot}3 \times 10^{-3}$ deg^{-1}, and zT_{mean} for room temperature is approximately equal to unity, individual specimens exhibiting even better properties, for high-temperature materials $z < 1 \times 10^{-3}$ deg^{-1} and zT_{mean} is considerably less than unity.

The basic expression for the thermal e.m.f. leads to semiconductors in which the chemical potential is close to zero. Therefore some degree of degeneracy of the free charge carriers is unavoidable.

When studying semiconductors in the degenerate state, Stil'bans and his colleagues succeeded in separating the effects of charge carrier velocity and thermal agitation of the lattice on the temperature dependence of mobility. They found that, in the investigated semiconductor systems, scattering is independent of the particle velocity, but depends on the lattice temperature.

Scattering increases in proportion to the absolute temperature at low temperatures, but starting with a certain temperature T_c the scattering increases at a faster rate. Therefore, both for PbTe and Bi_2Te_3 the mobility decreases with rise of temperature up to 150–200° K in proportion to $T^{-1{\cdot}5}$, and at higher temperatures it decreases as $T^{-2{\cdot}5}$ or even T^{-3}.

The decisive effect of the thermal conductivity on the performance of thermoelements makes it necessary to consider in detail the phenomena which may affect it. One of the factors, namely that of impurities in solid solutions, has already been considered. Let us now consider the effect of temperature. The phonon part of thermal conductivity κ_1 is, above the Debye temperature, as a rule inversely proportional to the absolute temperature. We have to add to this the thermal conductivity due to the free charge carriers. When the electrical conduction is due to impurities and proceeds through carriers of one sign only, this additional thermal conductivity is given by the Wiedemann–Franz law

$$\kappa_2 = A\frac{k^2}{e^2}T\sigma$$

where $A = 2+r$ in the absence of degeneracy, and $A = \pi^2/3$ for total degeneracy.

In modern thermoelectric materials κ_1 and κ_2 are quantities of the same order. Therefore a reduction in the value of κ_2 would be of great

importance. The factor A is 1·6 times greater for degenerate than for non-degenerate semiconductors, which may substantially affect the values of κ and z.

However, the onset of intrinsic conduction is of much greater importance. The diffusion of pairs of charge carriers of both signs transfers thermal energy in much greater quantities. This additional thermal conductivity was shown in 1940 by Davydov and Shmushkevitch to be equal to

$$\Delta\kappa_{\text{pair}} = \frac{k^2}{e^2} T \frac{\sigma_1 \sigma_2}{\sigma_1 + \sigma_2} \left(\frac{\Delta E_0}{2kT} + 2 + r \right)^2.$$

In intrinsic electrical conduction, the number of carriers of the two signs is equal. When their mobilities differ little from each other,

$$\frac{\sigma_1 \sigma_2}{\sigma_1 + \sigma_2} \doteq \tfrac{1}{4}(\sigma_1 + \sigma_2).$$

For $\Delta E_0 = 0{\cdot}5$ eV and $r = 0$, we have $\Delta\kappa_{\text{pair}} = 38k^2/e^2\, T\sigma$, whilst the thermal conductivity for the same electrical conductivity σ, according to the Wiedemann–Franz law would amount to

$$\kappa_2 = 2\frac{k^2}{e^2} T\sigma.$$

In other words, the transfer of heat by pairs of electrons and holes is greater by a factor of about 20 than that transferred by charge carriers of one sign only, for the same electrical conductivity.

It is important to remember that, even at temperatures when the current is transferred primarily by carriers of one sign, whilst carriers of the other sign participate in the current to a small extent, the effect of the latter on the thermal conductivity may become appreciable.

Thus, when the minority carrier current amounts to only 10% of the majority carrier current, the multiplier

$$\frac{\sigma_1 \sigma_2}{\sigma_1 + \sigma_2} = 0{\cdot}1(\sigma_1 + \sigma_2), \quad \text{and} \quad \Delta\kappa_{\text{pair}} = 8\kappa_2.$$

When $\sigma_1/\sigma_2 = 0{\cdot}025$, $\Delta\kappa_{\text{pair}} = \kappa_2$. Therefore the thermal conductivity due to free charge carriers is twice that which it would be with charge carriers of one sign only, although the participation of the carriers of the opposite sign increases the current by only 2·5%.

In Bi_2Te_3, with a forbidden gap width of 0·16 eV, intrinsic conduction begins to have a substantial effect on the thermal conductivity even below the room temperature. Thus, at a temperature of 250° K and a mobility of 300 cm^2/V sec, the intrinsic conductivity amounts to 100 $ohm^{-1}\ cm^{-1}$, whereas the impurity conductivity is usually of the order of 1000 $ohm^{-1}\ cm^{-1}$. Under these conditions the additional

thermal conductivity due to the diffusion of pairs becomes equal to about 20% of the thermal conductivity calculated from the Wiedemann–Franz law, although the concentration of minority carriers is only 5% that of the majority carriers.

At 300° K, $\Delta\kappa_{pair}$ amounts to 40% of κ_2 and at 400° K it leads to a doubling of the electronic thermal conductivity, although the minority carriers are responsible for only 15% of the total current. If the value of the electrical conductivity in Bi_2Te_3 were only 500 ohm^{-1} cm^{-1}, the effect of the intrinsic conductivity would be twice as great.

Thus, to delay as much as possible an appreciable increase of thermal conductivity due to the diffusion of pairs, it is necessary to increase the impurity conductivity of the semiconductor.

Another cause of additional thermal conduction at high temperatures may consist of heat transfer by radiant energy (infra-red radiation). Hentsel calculated that the heat transfer due to radiation is given by

$$\Delta\kappa_{rad} = \frac{16n^2}{3k} aT^3$$

where a is the black body radiation constant, n is the refractive index and k is the absorption coefficient of the material. For a forbidden gap width of ΔE_0 and in the absence of impurities and free charge carriers, the semiconductor is entirely transparent to all frequencies less than $\Delta E_0/h$. Under these conditions $\Delta\kappa_{rad}$ may be very large. With increasing temperature the multiplier aT^3 increases but at the same time the concentration of free charge carriers and the denominator k also increase.

Devyatkova and Moizhes showed that the combination of these effects leads to the peculiar temperature dependence of the additional thermal conductivity illustrated in Fig. 175, where the thermal resistivity of tellurium with different hole concentrations [10^{15} (curve I), 10^{17} (curve II) and 10^{18} (curve III)] is plotted. In the last curve the thermal resistance is proportional to the absolute temperature and is governed only by the value of κ_1; the additional heat transfer $\Delta\kappa_{rad}$ has no longer any effect owing to the high value of the absorption coefficient k. The different amounts of impurities in specimens I, II and III have a pronounced effect on the plot of $\log\sigma$ against $1000/T$ (Fig. 176).

For Bi_2Te_3 with a forbidden gap width of 0·16 eV and an impurity carrier concentration of 10^{19} cm^{-3}, the effect of infra-red radiation need not be considered.

In some semiconductors, such as PbTe, there seem to exist also other means of heat transfer due to the excited crystal states.

At high concentrations of free charge carriers the thermal conductivity decreases owing to the scattering of the phonons by electrons or holes. This effect becomes pronounced only at concentrations of the

order of 10^{20} cm^{-3} but at about 10^{22} cm^{-3} it virtually destroys κ_1. At low temperatures it is necessary to take into account the effect of the mutual dragging of phonons and electrons, which may raise the thermal conductivity of semiconductors by many times.

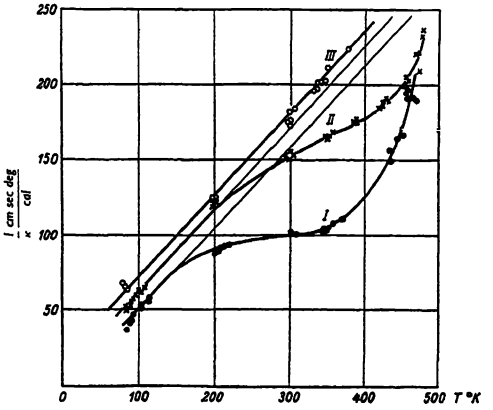


Fig. 175. Thermal resistivity of tellurium with different carrier (hole) concentrations. I, 10^{15}/cm^3; II, 10^{17}/cm^3; III, 10^{18}/cm^3.

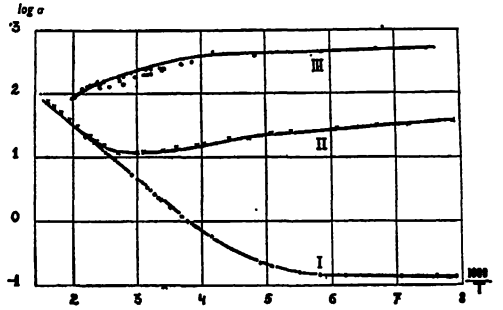


Fig. 176. Electrical conductivity of tellurium. (I, II, III, see Fig. 175.)

Finally, we shall also mention that, when reducing the thermal conductivity by forming solid solutions or by selecting crystals with a large number of defects, it is possible to prepare semiconductors with such low thermal conductivities (less than 0·001 cal/deg cm sec) that the mean free path length $\bar{\lambda}$ of phonons, calculated from the expression $\kappa = 1/3cv\bar{\lambda}$, is smaller than the lattice constant. For example, in some such materials we can take c as 0·4 cal/gm deg and v as $2{\cdot}5 \times 10^5$ cm/sec. Therefore, for $\kappa < 0{\cdot}001$ cal/deg cm sec we have $\bar{\lambda} < 3 \times 10^{-8}$ cm, whilst the lattice constant $a > 3 \times 10^{-8}$ cm.

When $\bar{\lambda} < a$ the concept of phonons is, of course, deprived of any sense. The mechanism and temperature dependence of the thermal conductivity acquire a different character and obey different laws.

References

1. O. V. Losev, "Luminosity of silicon carbide detectors; electrical conductivity of silicon carbide", *Phys. Zs.*, **32**, 692 (1931).
2. O. V. Losev, "Oscillating crystals", *Wir. World*, No. 271, 93 (1924).
3. See ref. 2 to Chap. 6.
4. E. I. Adirovich, *Nekotorye Voprosy Teorii Lyuminestsentsii Kristallov* (*Some Problems of the Theory of Luminescence of Crystals*). Gostekhizdat, Moscow–Leningrad, 1951.
5. I. V. Kurchatov and P. P. Kobeko, *Segnetoelektriki* (*Ferroelectrics*). Gostekhteoretizdat, Moscow–Leningrad, 1933.
6. B. M. Vul, *Segnetoelektriki* (*Ferroelectrics*), *Izd. Akad. Nauk SSSR.* Moscow, 1956.
7. G. A. Smolenskii, "A brief survey of some results of investigations of ferroelectrics", *Izv. Akad. Nauk SSSR, Ser. Fiz.*, **21**, 233 (1957).
8. A. V. Rzhanov, "Piezoeffect in barium titanate", *Zhur. Eksper. Teor. Fiz.*, **19**, 502 (1949).
9. K. S. Shifrin, "On the theory of electric properties of well conducting semiconductors", *J. of Phys. USSR*, **8**, 242 (1944).
10. A. F. Ioffe, *Semiconductors and Alloys.* Report to the General Assembly of the U.S.S.R. Academy of Sciences. Published reports, Moscow–Leningrad, 1948.
11. N. S. Kurnakov, *Vvedenie v Fiziko-Khimicheskii Analiz* (*Introduction to Physico-Chemical Analysis*). 4th edn. *Izd. Akad. Nauk SSSR*, Moscow–Leningrad, 1940.
12. G. B. Bokii, *Vvedenie v Kristallokhimiyu* (*Introduction to Crystal Chemistry*). Izd. Mosk. Univ., Moscow, 1954.
13. H. Y. Fan, "Valence semiconductors. Germanium and silicon", *Solid State Physics*, **1**, 283–365 (1954).
14. T. S. Moss, *Photoconductivity in the Elements.* London, 1952.
15. F. Stockmann, "Semiconductors", *Fortschr. Mineral.*, **33**, 1–111 (1954).
16. E. M. Conwell, "Properties of silicon and germanium", *Proc. IRE*, **40**, 1327 (1952).
17. R. Hafstadter, "Crystal counters", I, II, *Nucleonics*, **4**, 2 (1949).
18. F. C. Champion, "Electrical counting properties of diamonds", *Proc. Phys. Soc.*, **B.65**, 465 (1952).
19. E. A. Pearlstein and R. B. Sutton, "Mobility of electrons and holes in diamonds", *Phys. Rev.*, **79**, 5 (1950).
20. S. Mrosowski, "Semiconductivity and diamagnetism of polycrystalline graphite and condensed ring systems", *Phys. Rev.*, **85**, 609 (1952); W. Primak and L. H. Fuchs, "Electrical conductivity of natural graphite crystals", *Phys. Rev.*, **95**, 22 (1954).
21. G. L. Pearson and J. Bardeen, "Electrical properties of pure silicon and silicon alloys containing boron and phosphorus", *Phys. Rev.*, **75**, 865 (1949).
22. D. H. Chapin, C. S. Fuller and G. L. Pearson, "The Bell solar battery", *J. Appl. Phys.*, **25**, 676 (1954); *Bell Lab. Rec.*, **33**, 241 (1955).
23. V. K. Subashiev, *Poluprovodnikovye Preobrazovateli Solnechnoi Energii* (*Semiconductor Solar Energy Converters*). Leningrad, 1956.
24. A. R. Regel', A. I. Blum, and N. P. Mokrovskii, "Study of the electrical conductivity of semiconductors and intermetallic compounds in the solid and liquid states", *Izv. Akad. Nauk SSSR, Ser. Fiz.*, **16**, 139 (1952).

25. L. Landau and A. Kompaneets, "Deviation of semiconductors from Ohm's law in strong electric fields", *Phys. Zs. SU*, **6**, 163 (1934).
26. G. Dresselhaus, A. F. Kip, and C. Kittel, "Observation of cyclotron resonance in germanium crystals", *Phys. Rev.*, **92**, 827 (1953).
27. B. Lax, H. J. Zeiger, and R. N. Dexter, "Anisotropy of cyclotron resonance in germanium", *Physica*, **20**, 818 (1954).
28. K. Lark-Horovitz, "Nucleon-bombarded semiconductors", *Semiconducting Materials*, ed. H. K. Henisch. London, 47–69 (1952).
29. R. Cummerow, "Photovoltaic effect in p–n-junctions", *Phys. Rev.*, **95**, 16 (1954).
30. A. I. Blum and N. A. Goryanova, "Electrical properties of grey tin", *Dokl. Akad. Nauk SSSR*, **75**, 367 (1950).
31. G. Busch, J. Wieland and H. Zoller, "Electrical properties of grey tin", *Helv. Phys. Acta*, **24**, 49 (1951).
32. J. T. Kendall, "Electrical conduction in grey tin", *Phil. Mag.*, **45**, 141 (1954).
33. J. T. Kendall, "Electrical conduction in silicon carbide", *J. Chem. Phys.*, **21**, 821 (1953).
34. H. Welker and H. Weiss, "Group III–Group V compounds", *Solid State Physics*, **3**, 1–79 (1956).
35. L. Pincherle and J. M. Radcliffe, "Semiconducting intermetallic compounds", *Advances in Physics*, **5**, 271–322 (1956).
36. C. H. L. Goodman and R. W. Douglas, "New semiconducting compounds of diamond type structure", *Physica*, **20**, 1107 (1954).
37. V. P. Zhuze, V. M. Sergeeva and E. L. Shtrum, "Semiconductors with the structure of chalcopyrite", *Zhur. Tekh. Fiz.* (1957).
38. A. Hippel, "Structure and conductivity in the VIb-group of the periodic system", *J. Chem. Phys.*, **16**, 372 (1948).
39. H. W. Henkels, "Electrical properties of selenium", I, II, *J. Appl. Phys.*, **22**, 916, 1265 (1951); III, *ibid.*, **25**, 1 (1954).
40. F. Eckart, "Measurement of the conductivity of selenium", *Ann. d. Physik* (6), **14**, 233 (1954).
41. G. Borelius, F. Philstrand, J. Anderson and K. Gullberg, "Resistance of liquid and solidified selenium", *Arkiv für Mat., Astr. Fys.*, **30A**, 1 (1944).
42. B. T. Kolomiets, "Thallium sulphide photocells", *Izv. Akad. Nauk SSSR, Ser. Fiz.*, No. 5/6, 695 (1938).
43. V. E. Kosenko and E. G. Miselyuk, "Physical properties of silver sulphide photocells", *Zhur. Tekh. Fiz.*, **18**, 1369 (1948).
44. B. V. Kurchatov, "New types of rectifier", *Izv. Akad. Nauk SSSR, Ser. Fiz.*, No. 5/6, 679 (1938).
45. R. A. Smith, "Infra-red photoconductors", *Adv. Phys.*, **2**, 321 (1953).
46. E. D. Devyatkova, Yu. P. Maslakovets, T. S. Stavitskaya and L. S. Stil'bans, "Temperature dependence of carrier mobility in semiconductors", *Dokl. Akad. Nauk SSSR*, **84**, 681 (1952).
47. A. V. Ioffe and A. F. Ioffe, "Semiconductors in strong electric fields", *J. of Phys. USSR*, **2**, 283 (1940).
48. A. V. Ioffe and A. F. Ioffe, "The spectral distribution of the inner photoeffect in cuprous oxide", *Phys. Zs. SU*, **11**, 241 (1937).
49. I. K. Kikoin and M. M. Novikov, "A new photoelectric effect in cuprous oxide", *Zhur. Eksper. Teor. Fiz.*, **4**, 123 (1934).
50. V. E. Lashkarev, "The appearance of photo-e.m.f.'s in semiconductors", *Zhur. Eksper. Teor. Fiz.*, **18**, 917 (1948).

51. V. P. Zhuze and S. M. Ryvkin, "Photo-e.m.f. in cuprous oxide", *Izv. Akad. Nauk SSSR, Ser. Fiz.*, **18**, 93 (1952).
52. V. P. Zhuze and S. M. Ryvkin, "Exciton nature of light absorption and impurity conduction", *Dokl. Akad. Nauk SSSR*, **77**, 241 (1951).
53. E. F. Gross and N. A. Karryev, "Optical spectrum of excitons", *Dokl. Akad. Nauk SSSR*, **84**, 471 (1952).
54. E. F. Gross and B. P. Zakharchenya, "Excitons in cuprous oxide crystals", *Dokl. Akad. Nauk SSSR*, **90**, 745 (1953).
55. E. J. W. Verwey, *Oxidic Semiconductors, Semiconducting Materials*. London, 151–161 (1951).
56. S. I. Pekar, "The theory of F′-centres", *Zhur. Eksper. Teor. Fiz.*, **20**, 510 (1950).
57. S. I. Pekar and O. F. Tomasevich, "The theory of F′-centres", *Zhur. Eksper. Teor. Fiz.*, **21**, 1218 (1951).
58. B. I. Boltaks and V. P. Zhuze, "Semiconducting nature of some metal compounds", *Izv. Akad. Nauk SSSR, Ser. Fiz*, **16**, 155 (1952).
59. S. A. Vekshinskii, *Novyi Metod Metallograficheskogo Issledovaniya Splavov* (*A New Method for the Metallographic Study of Alloys*). Gostekhizdat, Moscow–Leningrad (1944).
60. I. V. Mochan, "Study of films of variable composition in the system Mg–Sb", *Dokl. Akad. Nauk SSSR*, **98**, 579 (1954).
61. E. D. Devyatkova, Yu. P. Maslakovets and L. S. Stil'bans, "The thermal conductivity and thermoelectric powers of the Sb–Zn system", *Zhur. Eksper. Teor. Fiz.*, **22**, 129 (1952).
62. A. R. Regel', "Dependence of the electrical conductivity of liquids on their structure", *Voprosy Teorii i Issledovaniya Poluprovodnikov i Protsessov Poluprovodnikovoi Metallurgii* (*Problems of the Theory and Investigation of Semiconductors and Processes of Semiconductor Metallurgy*). Izd. Akad. Nauk SSSR, 12–28 (1955).
63. A. T. Vartan'yan, "Temperature dependence of electrical conductivity in organic semiconductors", *Izv. Akad. Nauk SSSR, Ser. Fiz.*, **20**, 1541 (1956).

Conclusions

The first semiconductor rectifiers and photocells made their appearance at the end of the 1920's. Since that time semiconductors have been studied more and more extensively. A characteristic feature of this study during the past thirty years has been the close relationship between theory and experiment and between physics and engineering. As a result, a wide new field of science has been created, giving rise to a range of new industries.

The theoretical basis of this study was provided by the quantum theory of solids in the form adopted in the theory of metals by Sommerfeld, Bloch and Peierls. The periodic field of a crystal lattice leads to the band structure of the energy spectrum of electrons and the definition of the quasi-momentum of an electron. It also ensures the free motion of electrons and holes hampered only by lattice defects or thermal fluctuations, regarded as minor disturbances of the periodicity.

By averaging the motion of an electron over a length equal to many lattice spacings it is possible to ascribe to an electron in the conduction band a certain mean velocity and acceleration, which are analogous to the corresponding quantities in equations describing the motion of a free electron *in vacuo*; it is only necessary to replace the true mass in these equations by the effective mass.

The dependence of the energy of an electron on its quasi-momentum is assumed to be quadratic, and the form of the equal energy surfaces in momentum space to be spherical at low energies with a gradual transition to the shape of Brillouin zone with rise of energy.

The generally accepted theory of semiconductors has been successful in explaining many of their properties. Its framework has been found to be sufficiently wide to permit the inclusion of new facts such as the complex character of the energy bands and the tensor character of the effective mass. The theory in its present form cannot be used, however, to predict the nature or magnitude of these properties in a given material.

The mathematical apparatus and definitions of the quantities employed by the theory have been found to be at variance with the conditions arising in many semiconductors. Some assumptions are

suitable for certain types but unsuitable for other types of semiconductor (e.g. the assumption of the existence of a mean velocity and free path length). The lack of other basic assumptions (concerning the structure of chemical bonds) prevents the theory from covering some of the most important properties of semiconductors.

In spite of the great advances of the quantum theory there is an impression that this theory, so successful in solving the contradictions arising in the simple electron theory of metals, has far exceeded the limits of applicability of its principal assumptions in its universal employment for semiconductors.

The physical picture of free electrons with an effective mass m^* governed by Fermi statistics, describes the behaviour of metals well. There are, however, several reasons to doubt whether this picture is applicable in the case of semiconductors.

1. Semiconductors form an extensive class of materials with the electrical conductivity ranging from 10^4 to 10^{-10} ohm^{-1} cm^{-1}, the free carrier density ranging from 10^{10} to 10^{21} cm^{-3} and the mobility from 10^{-1} to 10^5 cm^2/V sec (at room temperature).

When going over from the mobility u to the corresponding carrier free path length, it is found that, at room temperature, they are interrelated by the expression

$$l \fallingdotseq 1 \times 10^{-8} \sqrt{\left(\frac{m^*}{m_0}\right)} u.$$

To large values of u there correspond small values of m^*. Confining ourselves to semiconductors for which $10^{-2} < m^*/m_0 < 10^2$, we find that the electron free path length lies within the limits from 10^{-9} to 10^{-4} cm. Let us compare these values with the interatomic distance a of about 5×10^{-8} cm and the wavelength λ of electrons moving freely at room temperature, 7×10^{-7} cm. It is found that for all semiconductors with a mobility u less than 5 cm^2/V sec, l is less than a; if u is less than 70 cm^2/V sec, l is less than λ.

Considering now temperatures of the order of 300° C we find that materials for which l is less than λ encompass all semiconductors with a room temperature mobility less than 200 cm^2/V sec.

Meanwhile free translational motion with a certain mean velocity v takes place only over a distance l. It is thus obvious that for all semiconductors with a mobility less than 100 cm^2/V sec the concept of velocity cannot be applied.

To avoid this dilemma one may assume a large value of the effective mass. However, in order to impart physical sense to the velocity of an electron, its acceleration and the effective mass itself, in the case of semiconductors with low mobility it would be necessary to attribute to

electrons and holes an effective mass of the order of a few thousand times the free electron mass. This would lead to the breakdown of the self-consistent field method, since the rates of attaining equilibrium among such heavy charge carriers and among atoms in the lattice would be the same order. In addition, the assumption that the effective masses are of the order of thousands m_0 is at variance with experimental values of the thermal e.m.f.s. As has been shown by Pekar, only in the case of polar states is this contradiction somewhat less pronounced.

In order to realize the theoretical importance of the existence of materials with u less than 100 cm^2/V sec, it must be noted that this class embraces the great majority of all known semiconductors (which includes, e.g. Se and TiO_2 with u less than 1 cm^2/V sec for which l is even less than a). In spite of a consistent search for materials with u greater than 1000 cm^2/V sec and therefore l much greater than λ, not more than about twenty such substances have so far been discovered.

2. The existence of semiconductor crystals with free path lengths of the order of interatomic distances makes doubtful the hypothesis that the disturbances of periodicity caused by thermal agitation are small. In fact, how would it be possible for small disturbances of the periodic field to prevent the free motion of charges with an energy within the limits of the conduction band? It must be conceded that thermal agitation changes the energies and frequencies of vibrations of electrons in neighbouring atoms to such an extent that their motion in the conduction band over distances equal to several interatomic spacings becomes impossible.

One cannot speak of the mean velocity of such electrons or define their orbits. Displacement of such electrons in the course of diffusion or under the effect of an external field may be thought of as the residence of an electron for a certain period of time within the limits of one lattice site and subsequent transition to one of the neighbouring sites, with a certain statistical probability which decreases with the intensity of thermal agitation. It should be noted that Frenkel' based his theory of metals, dating back to 1928, on exactly the same assumptions. Frenkel's theory is more applicable to semiconductors with low mobilities, whereas Sommerfeld's theory and the kinetic theory of gases are more closely related to high values of the mobility.

The consideration of thermal conduction leads to the same conclusion, viz. that the disturbances produced by thermal agitation are large. The small free path length of phonons in semiconducting crystals implies a high degree of anharmonicity of the thermal vibrations; the atomic displacements extend far beyond the limits of Hooke's law.

3. Regel's experiments have shown that the energy spectrum changes little in the transition from the solid into the liquid state, provided that

the short-range order (i.e. coordination number) does not also change during melting. This will be easily understood by regarding the bands as resulting from the splitting of discrete energy levels of atoms as they approach each other until the wave functions of neighbouring atoms overlap. On the other hand, the derivation of the band structure from the periodic properties of crystals ceases to carry conviction when considering semiconductors with low mobilities, and even more so in the case of liquid and vitreous semiconductors.

4. We are concerned here with another facet of the problem of the theory of semiconductors. Experiment shows that all the basic properties of a semiconductor, including its band structure, are governed by the short-range order, i.e. chemical bonds within a single crystal cell. Saturation of the chemical bonds in molecules leads to a filled valence band; when a solid forms from such molecules it becomes a semiconductor. The close vicinity of an adjacent crystal cell introduces only other boundary conditions than those for an isolated molecule. Translational symmetry therefore plays the role of providing boundary conditions, which make it possible to use a convenient mathematical apparatus. But the chemical bonds and the band structure following from them, which are the basic factors governing the properties of any given semiconductor, remain outside the theory.

In semiconductors, as opposed to metals, the electrons with energies close to the band edges are of the greatest importance. In recent years, owing to the study of cyclotron resonance and the effect of a magnetic field on the resistivity, the great importance of the complex structure of energy bands has been established for such apparently simple crystals as germanium and silicon. One cannot maintain any longer that it is sufficient to describe a semiconductor by specifying the width of the forbidden gap and the scalar values of the effective masses of holes and electrons.

It is necessary to develop methods which would make it possible to relate the properties of semiconductors to the structure of the electron cloud in a cell. Samoilovich has already succeeded in obtaining interesting results with such an approach to the properties of free holes and electrons in germanium.

5. Comparison of the optical absorption spectra of such crystals as Ge and Si with their energy band systems shows that optical transitions requiring the assistance of phonons are the most probable.

The temperature dependence of the mobility compels us to think that at elevated temperatures transitions of electrons with the participation of several phonons acquire a high probability.

Electron transitions between the individual valleys forming the band structure, the anisotropic behaviour of layer-like crystals and the

existence of spiral, stratified and straight chain structures, all render the energy or scalar approach to phenomena in semiconductors, which has been employed for many years, entirely inadequate; this is so even for crystals belonging to the cubic system.

It is now quite trivial to express the view that neither in an electrical nor in a thermal field is it possible to consider the motion of electrons and phonons assuming that each portion of their path is in statistical equilibrium. This is clearly demonstrated by the drag effects.

Finally, there is no doubt that the energy state spectrum is not confined to the valence and conduction bands. The right of citizenship has already been granted to polarons, excitons, plasmons, impurity bands, surface states and a whole series of defects, deep and shallow impurity levels, traps with long and short lifetimes, etc.

All this compels us to reconsider the problem of interaction between electrons and photons with phonons in a solid.

Modern theories of semiconductors do, of course, take account, to a certain degree, of the nature of the chemical bonds, the symmetrical or polar distribution of charge carriers within a cell; when the distribution is symmetrical account is taken only of the acoustical branch of the thermal vibrational spectrum, but when the charge disposition is polar the optical branch becomes predominant. The scattering of electrons in these two cases is also estimated in different ways.

It is evident, however, that such general considerations are inadequate and that they often lead to conclusions at variance with experiment (e.g. when considering the temperature dependence of the mobility).

All that has been said shows that the theory of semiconductors is undergoing a transition to a new stage, where the concepts carried over from the theory of metals no longer hold. This also applies to the mathematical methods employing the kinetic equation and small perturbations. The chemistry of the molecule is becoming one of the starting-points for the physics of semiconductors. Theory is ceasing to play the role of a formal external frame filled by experiment. One must learn to take account of the physico-chemical nature of any given semiconductor.

An assessment of the present state of the science of semiconductors leads us to the following conclusions:

I. The presence in liquid and amorphous semiconductors of a band structure in the energy spectrum of electrons, and the hole and electronic conduction observed in such semiconductors, prove the following hypotheses:

(a) The band structure stems from more general ideas than the periodicity of the electrical field in crystals, namely from the discrete character of the spectra of the atoms and molecules forming the crystal.

(b) In an amorphous solid, in which the energy levels of neighbouring atoms differ, movements of charge, occurring even in the conduction band, cannot be regarded as movements with a free path length extending over a large number of atomic distances. Consequently it is not possible to apply to the motion of carriers the concepts of average velocities and accelerations or to describe them by two-dimensional waves with a wavelength λ. Instead, it will be necessary to develop a theory of electrical conduction and diffusion on the basis of electrons localized for a certain length of time within the limits of an individual cell, but capable of passing, with a certain probability, from one cell into another, in a manner resembling the motion of impurity ions in an electrolyte.

(c) At the same time, the concepts of the energy spectrum of electrons in a solid as a whole and of wave vectors should be retained and be determined by the macroscopic value of the average density of the solid and the short-range order, which changes little in the transition from the crystalline to the amorphous state.

II. The low mobility of charge carriers in the majority of semiconductors leads us to the following conclusions:

(a) The concept of mean velocities of motion of electrons, extending over the free path length, is devoid of meaning in the case of most semiconductors which have low mobilities (less than 100 cm^2/V sec at room temperature) as well as liquid and amorphous semiconductors.

(b) The movement of charge carriers with an energy corresponding to the conduction band, consists in such crystals, as well as in amorphous semiconductors, of transitions from one lattice site to another with a probability decreasing with increasing intensity of thermal agitation, and not of individual free paths interrupted by disturbances of the periodicity.

The smaller the mobility of the "free" charge carriers the greater the part played, in addition to that of the scattering by fluctuations of the electrical field, by the decrease of the probability of transitions among neighbouring sites.

(c) Disturbances of lattice periodicity due to thermal agitation cannot be regarded as small disturbances.

It is interesting to compare the scale of disturbances of the lattice periodicity, lying at the root of the low mobility of electrons, with the amplitudes of the thermal vibrations, producing anharmonicity, responsible for thermal resistance and thermal expansion.

(d) Experiments carried out by Airapetyants and Stavitskaya demonstrate that the mobility of electrons is governed mainly by the configuration of positive ions, whereas the motion of holes is related to negative ions. These ions are separated by distances of the order of a few angstroms, whereas the wavelengths of both electrons and holes extend over tens of such distances. These facts cannot be explained by the motion of two-dimensional waves scattered on isolated defects; it is necessary either to think in terms of modulated waves or to introduce semi-classical concepts with regard to the paths of motion of individual charge carriers.

The above conclusions make the theory regarding thermal agitation as giving small disturbances of the periodicity, and the motion of electrons in the lattice as "free" motion, differing only from that *in vacuo* as far as the value of the effective mass is concerned, inapplicable to semiconductors with low mobilities and to amorphous semiconductors.

The theory of energy bands and kinetic processes in semiconductors should be developed on a wider basis and should embrace not only semiconductors with properties approaching those of metallic crystals, but also all other solids with all kinds of defects of periodicity, extending to liquids and amorphous solids.

III. In semiconductors with high values of the mobility, the free path is so long that at low temperatures it is possible to observe cyclotron resonance. Its study, as well as an investigation of the effect of a magnetic field on the electrical conductivity, lead to the following conclusions:

(a) The structure of energy bands is much more complex than has been hitherto assumed; energy minima may be situated at several points in momentum space, whilst charge carriers with different mobilities are found within the same energy limits. The energy minimum does not always coincide with the zero value of the wave vector.

(b) The effective mass of every type of charge carrier should be described by a tensor, even in the case of crystals belonging to the regular system.

(c) The nature of motion of a charge carrier in the conduction band is completely different when its energy changes within the boundaries of a single valley of the band structure from that when it is from one valley to another.

Although the complex structure of energy bands is characteristic of all solids, it remained in the shade so long as physics concerned itself with metals, making its presence felt only through the de Haas–van Alphen phenomenon and the complex nature of the effect of a magnetic field on the electrical conductivity of metals. Since kinetic phenomena in metals are governed by electrons within a small range of energies

close to the Fermi level, quite remote from the lower edge of the conduction band, the structure of the latter is found to have virtually no effect.

On the other hand, in semiconductors in which the free charge carriers are in the immediate vicinity of the band edge, the band structure comes to the fore. It should, therefore, be experimentally investigated and provided with a theoretical basis.

The problem of small free path lengths is in a similar state. In metals, the kinetic energies of electrons are hundreds of times higher than in non-degenerate and even degenerate semiconductors, whilst the free path length l is much larger. Therefore metals usually satisfy the condition that l should be much greater than λ.

Incidentally, the problem of the existence of liquid conductors side by side with crystalline ones applies to the same extent to metals.

IV. The difference between semiconductors and metals, which in the band picture is determined by the filling of valence band levels, owes its origin to the saturation of chemical bonds in the substance. Mooser and Pearson, basing themselves on Pauling's theory, have formulated in a simple algebraic expression the condition of filling of the valence band. Mooser and Pearson's equation relates the semiconducting character of the substance to the nature of chemical bonds between constituent atoms. Going further than that, it may be maintained that all the properties of semiconductors—the forbidden gap width, form of energy levels, the effective mass, mobility, thermal conductivity, photoconductivity, etc.—are determined by the chemical bonds forming the basic cell of the solid, whether crystalline or amorphous.

Chemical bonds in turn manifest themselves in the distribution of electron density within a cell. Therefore, the structure of the electron cloud within the confines of a cell may be regarded as the primary characteristic of a semiconductor.

This postulate is clearly illustrated by the results of measurements of the thermal conductivity of potassium chloride crystals coloured by F-centres, carried out by Devyatkova and Stil'bans. It was found that $4{\cdot}8 \times 10^{18}$ cm^{-3} F-centres reduced the thermal conductivity κ from $18{\cdot}6 \times 10^{-3}$ to 16×10^{-3} cal/deg cm sec. A comparison of the free path lengths of phonons leads us to the conclusion that a single F-centre possesses a phonon-scattering cross-section equal to 25 crystal cell cross-sections. Therefore a single F-centre distorts around itself a crystal volume equal to 125 cells to such an extent that this volume scatters plane acoustic waves with corresponding wavelengths.

However, the only difference between an F-centre and a normal site occupied by a negative chlorine ion is the fact that, instead of the electron density being distributed throughout the cell, there is a single central charge.

Both the physical and chemical properties of semiconductors follow from the character of the valence bonds in the lattice.

(a) For this reason it is necessary to develop methods which would permit us to determine the properties of semiconductors starting with the structure of the electron cloud inside the crystal cell, and to use translational symmetry for the determination of boundary conditions for the electron equilibrium in an isolated cell.

(b) Since the density distribution of the electric charge in cells may be determined experimentally, we are faced with the problem of establishing a relationship between the geometric shape of the cloud and the form of the energy bands.

(c) A regular dependence between molecular bonds and the energy spectrum of electrons in a crystal may also be expected.

V. Optical absorption spectra of many semiconductors exhibit energy states unrelated to electrical conduction, which require a smaller excitation energy than the transition into the conduction band. The existence of such states was predicted in 1931 by Frenkel', who called them excitons. Subsequently, Mott and other investigators identified excitons with the states of combined electrons and holes, resembling positronium.

A study of thermal conductivity has led us to the concept of excitation states, created by thermal agitation and diffusing in a semiconductor, which emit their absorbed energy in their passage from the hot to the cold portions of the body. In the special case of lead telluride and lead selenide it was found possible to correlate the energy transported by these states with the energy of the optical absorption quanta.

There are, however, substantial differences between the excitation states apparent in thermal conductivity measurements, and excitons. Whereas excitons of the positronium type should possess an energy close to the forbidden gap width, the thermal conductivity may be increased only by the contribution of excitation states with an energy less than half the forbidden gap width. Elementary considerations show that such excitations cannot form positronium.

It seems that Mott's exciton, resembling positronium, will have to be regarded as a special case of the excitations which are possible in a solid and capable of diffusing in it. All rearrangements of energy of electrons in a cell can lead to the appearance of excitons. In particular, optically inactive excitation states may exist, which lead to an increased thermal conductivity, without exhibiting an absorption band in the infra-red spectrum.

It is natural to expect that exciton-type excitations will also have an effect on the electron mobility and on the thermal e.m.f. in semiconductors.

It is essential to study the influence of excitons on the magnetic properties of semiconductors and the possibility of the existence of excitons with uncompensated spins.

VI. The facts relating to the interaction of electrons and photons with phonons necessitate a review of the question of selection rules, multi-phonon transitions and the probability of transitions of electrons between states of a solid, and also, of the statistical equilibrium in the presence of an electric field or a temperature gradient.

VII. The methods of non-Bragg scattering of X-rays, scattering of X-rays at small angles and other means of investigation have revealed the presence of numerous defects both in pure and, in particular, in impurity semiconductors. These methods provide us with the means of (a) studying the number, dimensions and shape of distortions introduced by impurities and irradiation; (b) studying the structure of subtractional, interstitial and substitutional solid solutions, and the dimensions of the distorted portions of a crystal in each of these cases, and establishing the limits of solubility of impurities.

VIII. Precise methods of investigation of the surface properties of semiconductors provide the means for studying the catalytic properties of semiconductors.

The advance of the study of semiconductors in the directions outlined above may be expected to create a new link between the physics and chemistry of solids.

Subject Index

Absorption, optical, 77, 345
Acceptors, 123–124, 155–156
Activation energy, 152
Aluminium antimonide, 350, 374–375
Aluminium arsenide, 350
Aluminium oxide, 349, 393
Ambipolar diffusion, 183
Amorphous solids, 95–96
Antimony, 380
Antimony-caesium photocathodes, 223
Antimony sulphide, 312
Antimony telluride, 407, 409
Arsenic, 380
Atomic lattices, 353–376

Barium oxide, 312, 350
Barium titanate, 350, 395
Barrier layer, 82–84
Bismuth, 250–251
Bismuth sulphide, 312
Bismuth telluride, 285–286, 407–410
Bismuth telluride alloys, 402, 410, 412–416
Boron, 376
Boron nitride, 354
Breakdown, 79–82
Bridgman effect, 292, 305
Brillouin zones, 104–105
Brownian motion, 67

Cadmium antimonide, 398
Cadmium oxide, 312
Cadmium sulphide, 350, 381
Cadmium telluride, 386
Caesium antimonide, 350, 398
Caesium oxide, 233
Caesium sulphide, 350
Calcite, 20–21
Cathodic sputtering 241
Chalcopyrite, 375
Chemical potential, 116–118, 186, 208
Chromium silicide, 414
Classification of semiconductors, 352–353
Cobalt antimonide, 414
Concentration of electrons, 68–69, 79–82, 114–115, 184–187, 332–333
Concentration of ions, 9–11
Conduction band, 225
Conduction, ionic, 1
Conductivity (see electrical conductivity and thermal conductivity)
Contact electrification, 242–243
Contact potential, 53, 83, 192, 228–231, 342–344
Contact resistance, 189–196
Coordination number, 143–145
Copper bromide, 19
Copper chloride, 19
Copper iodide, 19
Crystals, 95, 137–156
Cuprous oxide, 81, 83, 121–123, 129, 132, 150, 178–179, 194–195, 198, 200, 311–312, 326–328, 349, 386–391
Cuprous sulphide, 327, 381–383
Cyclotron resonance, 262–266, 333

Debye temperature, 159–161
Degeneracy, 41–42, 176
Diamagnetic resonance (see cyclotron resonance)
Diamagnetism, 257–258
Diamond, 21, 139–141, 350, 353
Dielectric constant, 89–90, 346
Diffusion, 206, 208–209
Diffusion coefficient, 67, 302
Diffusion length, 67, 335–336
Dislocations, 16, 149
Dissociation energy, 5
Donors, 123–124, 155–156

Effective mass, 100–104, 106–111, 118, 174–175, 265–266, 332–333
Einstein's equation, 67, 182, 303
Electric fields, strong, 79
Electrical conductivity of metals, 30–32
Electrical conductivity of semiconductors, 68–69, 328–330
Electrolytes, solid, 1–23, 57, 60
Electron theory of metals, 32–33, 111–112
Electronic semiconductors, 59–86
Electrons, 24, 93
Electrons, conduction, 91, 96–109
Emission of electrons, 221–228, 240–241
Energy bands, 87–91, 367–372
Energy conversion, 351
Energy distribution, 38–40
Energy levels, 129–132
Energy of electrons, 62
Ettingshausen effect, 255
Excitons, 75, 128–129

F-centres, 234
Faraday's law, 2–3
Fermi–Dirac statistics, 34, 309
Fermi level, 38–40, 47, 53, 57, 113
Ferrites, 270–271
Ferroelectricity, 350
Ferromagnetism, 266–270
Forbidden band, 91, 102–103, 333–335
Free path, 173–184
Frenkel' defects, 6–9, 149

Gallium antimonide, 350, 374–375
Gallium arsenide, 375
Galvanomagnetic effects, 255
Germanium, 67, 111, 154, 156, 173, 198, 207, 209, 212–214, 249, 265, 312, 331, 349, 361–372, 400
Germanium-silicon alloys, 414
Glasses, 95, 275
Graphite, 354
Grey tin, 144, 372–374

Hall coefficient, 24, 331
Hall effect, 4, 24–25, 63, 110–111, 185, 245–249
Heat capacity (see specific heat)
Holes, 65, 104–109
Hydrogen bond, 142

Impurities, 16–17, 66, 69–72, 92, 151–156, 333, 370–371
Impurity band, 183
Impurity levels, 120–125, 333–335
Indium antimonide, 111, 249, 285, 350, 374–375
Indium arsenide, 375
Indium phosphide, 375
Indium telluride, 386
Intrinsic conduction, 185
Ionic bonds, 139–140
Ionic conduction, 204–205
Ionic crystals, 59, 180–181, 395–396
Ions, 2, 93
Ions, interstitial, 6

Lattice, 137–167
Lattice defects, 15
Layer lattices, 146
Lead chloride, 17
Lead selenide, 282, 382–385, 404
Lead sulphide, 178–179, 272, 304, 312, 350, 382–385
Lead telluride, 273, 282, 284–286, 304, 350, 382–385, 404–405
Lead telluride alloys, 402, 404–407, 411–412
Lifetime, 335–336, 345
Liquid metals, 95
Liquid semiconductors, 399
Lithium, 57
Luminescence, 75, 349–350

Maggi–Righi–Leduc effect, 255
Magnesium antimonide, 397–398
Magnesium germanide, 398
Magnesium silicide, 398
Magnesium stannide, 398
Magnetic properties, 55–56
Magnetic susceptibility, 56
Magneto-resistance effect, 249–256
Manganese silicide, 414
Manganese telluride, 179
Maxwell–Boltzmann law, 29
Mean free path (see free path)
Measurements, 326–348
Mercury selenide, 111, 173, 249, 375
Metal alloys, 396–398
Metallic bond, 59
Metals, 24–58, 61

Mobility, electronic, 24–27, 61, 66–67, 78–82, 109–111, 173–184, 331–332
Mobility, ionic, 2, 13–14
Modes of vibration, 156
Molecular crystals, 59, 376–380
Molybdenum disulphide, 312
Molybdenum oxide, 392

Nernst effect, 255, 331
Nernst–Ettingshausen effect, 255
Nickel oxide, 312
Noble gases, 147
Nuclear radiation, 221

Ohm's law, 24, 79, 110
Order, lattice, 136–139
Organic semiconductors, 400–401
Oxides, 386–395

p-n junctions, 205–214
Paramagnetism, 258–262
Peltier coefficient, 293
Peltier effect, 56, 290
Periodic field, 99–100
Phonon-drag effect, 309–310
Phosphors, 350, 380
Photoconduction, 72–79
Photoconductivity, 344–346
Photomultiplier, 239
Photons, 75, 91, 94, 220
Photovoltaic effect, 79, 84, 219–221, 231–234
Piezoelectricity, 351
Pisarenko's expression, 307
Polarization, 20-21, 330
Polarons, 127, 396
Positive holes (see holes)
Potassium chloride, 17, 129

Quantum states, 43–45, 86–91, 112–114
Quantum statistics, 34, 43
Quantum theory, 84–85, 87–135
Quartz, 15, 16, 21, 137, 275

Radiation, 287
Radiation detection, 349–350
Recombination, 207
Rectification, 196–203
Rectifiers, 214–219, 349
Refractive index, 345
Residual rays, 165
Resonance scattering, 177
Righi–Leduc effect, 255
Rock salt, 8, 17, 19, 21, 75, 88–89, 395–396
Rutile (see titanium oxide)

Saltpetre, 20
Scattering of electrons, 167–173
Scattering of phonons, 276–283
Schottky defects, 8–9, 149
Secondary emission, 237–240
Seebeck effect, 56, 288
Seignette salt, 350
Selection rules, 78
Selenium, 95, 173, 198, 200, 349, 376–379
Semiconductors (see electrolytes, solid and electronic semiconductors)
Silicon, 125, 154, 156, 173, 198, 214, 349, 354–361, 370–371
Silicon carbide, 199, 215, 312, 349, 374
Silver bromide, 8, 15,
Silver chloride, 9, 15
Silver sulphide, 4, 19
Silver telluride, 386
Sodium fluoride, 17
Solar generation, 220
Solid solutions, 410–416
Specific heat, 42–43, 159–161, 337–342
Spinel, 393–394
Stannite, 375
Sulphur, 376
Surface layers, 188–189
Surface states, 125–127

Tellurium, 156, 380
Thallium sulphide, 312, 381
Thermal conductivity, lattice, 274
Thermal conductivity of metals, 27–30
Thermal conductivity of semiconductors, 271–287, 337–342, 416–419
Thermal e.m.f. (see thermoelectric power)
Thermal vibrations, 156–166
Thermionic emission, 51–55, 234–236, 343–344, 350
Thermistors, 349
Thermoelectric cooling, 318–321
Thermoelectric figure of merit, 315–317
Thermoelectric generators, 313–317
Thermoelectric heating, 321–322

Thermoelectric phenomena, 56–57, 189, 288–321
Thermoelectric power, 111, 293, 305–313, 336–337
Thermoelements, 220, 401–419
Thermomagnetic effects, 255
Thomson coefficient, 293
Thomson effect, 291
Thomson heat, 315
Thorium oxide, 350
Time-dependent phenomena, 203–205
Tin (see grey tin)
Tin sulphide, 312, 381
Titanium oxide, 83, 111, 173, 198, 312, 327, 391–392
Transfer numbers, 3
Transistors, 214–219
Tungsten oxide, 83, 111, 173, 310, 392

Uranium oxide, 83, 327, 349, 391

Valence band, 225
Valence bonds, 59, 139–140, 353
Van der Waals forces, 59, 139, 141
Vanadium oxide, 83, 311
Velocity of electrons, 49–51, 79–82, 96–97
Vibration spectrum, 162–163

Waves, electronic, 61
Waves, thermal, 61
Wiedemann–Franz law, 29, 272, 283–284
Work function, 51, 225

X-ray spectra, 45–49, 118–119

Young's modulus, 142–143

Zinc antimonide, 397–398, 402–404
Zinc oxide, 83, 132, 327
Zinc sulphide, 19, 327
Zone refining, 154